По океану дискретной математики

Ю. А. ЗУЕВ

Том 1

ОТ ПЕРЕЧИСЛИТЕЛЬНОЙ КОМБИНАТОРИКИ ДО СОВРЕМЕННОЙ КРИПТОГРАФИИ

Основные структуры

Методы перечисления

Булевы функции

URSS

Ю. А. ЗУЕВ
По океану дискретной математики
ОТ ПЕРЕЧИСЛИТЕЛЬНОЙ КОМБИНАТОРИКИ ДО СОВРЕМЕННОЙ КРИПТОГРАФИИ
ТОМ 1
URSS

Ю. А. Зуев

ПО ОКЕАНУ ДИСКРЕТНОЙ МАТЕМАТИКИ

От перечислительной комбинаторики до современной криптографии

Том 1

ОСНОВНЫЕ СТРУКТУРЫ

•

МЕТОДЫ ПЕРЕЧИСЛЕНИЯ

•

БУЛЕВЫ ФУНКЦИИ

URSS
МОСКВА

ББК 22.176

Зуев Юрий Анатольевич

По океану дискретной математики: От перечислительной комбинаторики до современной криптографии. Т. 1: Основные структуры. Методы перечисления. Булевы функции. — М.: Книжный дом «ЛИБРОКОМ», 2012. — 274 с.

Содержание настоящей книги охватывает вузовский курс дискретной математики, включая перечислительную комбинаторику, булевы функции, графы, алгоритмы, помехоустойчивое кодирование и криптографию, а также ряд дополнительных тем. Принцип построения «от простого — к сложному» делает начальные разделы каждой главы доступными для старшеклассника, а заключительные — ценными для аспиранта. Для самостоятельного решения предлагается большое число задач различной сложности, снабженных ответами и указаниями. В книге рассказывается также об истории математических открытий и формулируются открытые проблемы дискретной математики.

Книга состоит из двух томов. В первом томе даются основные идеи и понятия дискретной математики, изучаются теория и методы перечисления, булевы функции. Второй том, посвященный графам, алгоритмам в дискретной математике, теории кодирования, выходит одновременно с первым в нашем издательстве.

Написанная доступным языком, в яркой форме и с многочисленными примерами, книга будет полезна широкому кругу читателей, желающих познакомиться с основами дискретной математики.

Издательство «Книжный дом "ЛИБРОКОМ"».
117335, Москва, Нахимовский пр-т, 56.
Формат 60×90/16. Печ. л. 17,125. Зак. № ЖР-25.

Отпечатано в ООО «ЛЕНАНД».
117312, Москва, пр-т Шестидесятилетия Октября, 11А, стр. 11.

ISBN 978–5–397–02572–0

НАУЧНАЯ И УЧЕБНАЯ ЛИТЕРАТУРА

E-mail: URSS@URSS.ru
Каталог изданий в Интернете:
http://URSS.ru
Тел./факс (многоканальный):
+ 7 (499) 724–25–45
URSS

10237 ID 122779

Напутствие от старого морского волка

Взяв в руки эту книгу, Вы, уважаемый читатель, получаете возможность совершить увлекательное путешествие «По океану дискретной математики», в пяти морях которого Вас ждут удивительные встречи и открытия. Каждое море встретит Вас семью футами под килем, затем появятся мели и рифы, но через них и лежит путь к настоящим открытиям. Однако Вы сами являетесь здесь капитаном и решаете, как далеко заплывать и когда менять курс, не ложась, разумеется, в дрейф. Прилагаемая лоция с отмеченными на ней маяками и стрелка компаса Ваших устремлений помогут выбрать маршрут.

Если Вы ещё не морской волк, то, пройдя пять морей, несомненно, станете им. Но в этом случае перед выходом в открытое море стоит совершить небольшое каботажное плавание, которое Вас также не разочарует. Оно привьёт Вам вкус к морю, даст необходимые в морском деле навыки и оставит неизгладимый след в Вашей памяти, даже если Вы не собираетесь стать моряком. А если Вы решили связать свою жизнь с морем, то, проборoздив пять морей, заглянете и в два залива, плавание по которым требует уже солидной оснастки.

Итак, в путь! Выбрать якоря! Поднять паруса! Налечь на штурвал! Круче к ветру! Так держать!.. И выдержать выбранный курс наперекор рёву ветра и ярости волн!!!

Никто не обнимет необъятного.

Козьма Прутков

Превосходна та книга, в которой всё, от начала до конца, изложено в стройном порядке, ничто не упущено, ничто не добавлено некстати, в которой соблюдена соразмерность отдельных частей, которая всё разъясняет и в которой всё обосновано.

Джироламо Кардано

Оглавление

Предисловие

Стремительно развивавшаяся во второй половине XX века дискретная математика заняла в конце его важное место в общем курсе математической подготовки студентов университетов и технических вузов. Владение её элементами стало обязательной составной частью математического образования инженеров, экономистов, специалистов по вычислительной технике. Причины возросшего интереса к дискретной математике понятны. В эпоху, когда стремительно нарастает использование вычислительной техники как в теоретических исследованиях, так и в прикладных задачах, а компьютер вошёл во все сферы человеческой деятельности, само человеческое общество из индустриального становится информационным. Развитие информационных технологий, решение проблем, связанных с автоматической обработкой информации и принятием решений, стало приоритетным направлением развития науки. При этом возникает огромное число задач дискретного характера.

Истоки дискретной математики

Задачи дискретной математики, которыми в XVIII веке занимался гениальный Леонард Эйлер, в течение двухсот лет оставались единичными примерами, носившими характер развлекательных головоломок, привлекших к себе внимание, в значительной мере, благодаря имени великого математика. В XIX веке Кирхгоф, строя теорию электрических цепей, фактически работал с графами, но тогда это понятие оказалось еще недостаточно востребованным, чтобы выделить его из приложения. Даже книга Кёнига, вышедшая в Германии в 1936 году и впервые осветившая теорию графов как самостоятельную математическую дисциплину, была по достоинству оценена лишь двумя десятилетиями позже, после того, как в 1950 году вышло её американское издание.

Создание дифференциального исчисления в конце XVII века позволило от господствовавшей в античном мире статики перейти к динамике физических процессов. XIX век стал триумфом непрерывной математики, описавшей дифференциальными уравнениями поведение материи и электромагнитного поля, рассматриваемых как непрерывные среды. Принцип непрерывности, согласно которому физическая система переходит из одного состояния в другое путём бесконечно малых изменений, был отчётливо сформулирован Лейбницем: *«Природа никогда не делает скачков»*. И в соответствии с этим дифференциальные уравнения надолго стали единственным математическим инструментом моделирования физической реальности.

В XX веке выяснилось, однако, что природа скачки делает, и идея дискретности всё настойчивее стала пробивать себе дорогу. Физика, столкнувшись с квантовыми явлениями, стала изучать атомы и элементарные частицы, возникла квантовая теория. Биология перешла от дарвинской теории непрерывного развития видов к изучению генов и кодов наследственности.

Кодирование и криптография ➡

Вторая мировая война усилила интерес к быстрой обработке информации и анализу данных и стимулировала создание компьютеров, первоначально использовавшихся лишь в качестве мощных калькуляторов, но вскоре совершивших переворот в области информационных технологий. С появлением компьютеров цифровое кодирование информации вытеснило запись её на аналоговых носителях, таких как музыкальные пластинки, магнитные ленты и фотокарточки. Дефекты старых аудио- и видеозаписей, шипение, треск и помехи, возникающие в результате перезаписи, а также вследствие физической изношенности носителей, не угрожают записям, использующим цифровые технологии. Сделанные сегодня, эти записи и через тысячу лет сохранят при воспроизведении такую же чистоту звука и изображения.

Криптография, возраст которой насчитывает не одну тысячу лет, но основными пользователями которой на протяжении столетий были лишь разведчики, военные и дипломаты, пережила второе рождение. Основанные на компьютерных технологиях новые удобные способы защиты информации от несанкционированного доступа способствуют всё более широкому проникновению криптографии во все сферы нашей жизни. Уже сейчас она нашла себе широкое применение в банковском деле, где наряду с секретностью финансовых распоряжений возникло понятие «электронной подписи» — цифрового сообщения, подтверждающего аутентичность финансового распоряжения.

Дискретность и непрерывность ➡

Сегодня даже в тех физических задачах, где непрерывные модели сохраняют своё значение и для описания физических процессов используются дифференциальные уравнения в частных производных, при практическом их решении на компьютере используется метод сеток и производные заменяются конечными разностями.

С развитием компьютерных технологий во второй половине XX века дискретная математика в мгновение ока из Золушки превратилась в блистательную принцессу. Но хотя компьютер и дал мощный толчок развитию дискретной математики, она родилась задолго до его появления. Две ветви математики, *дискретная* и *непрерывная*, возникли, когда люди стали считать и измерять. Сегодня можно сказать, что дискретность и непрерывность являются такими же неразрывно связанными и взаимно дополняющими друг друга основными концепциями в математике как корпускулярная и волновая теории света в физике.

Теория чисел — это древнейшая ветвь дискретной математики. Предметом её рассмотрения, однако, является уникальное дискретное множество, созданное, по выражению Кронекера, самим Господом Богом — множе-

ство натуральных чисел. Для его изучения она использует специфические, наработанные веками методы, которые применяются к задачам, зачастую также имеющим многовековую историю.

Современная дискретная математика рассматривает дискретные структуры самой различной природы и разрабатывает общие методы работы с подобными структурами. Всё большую роль в ней играют проблемы, связанные с алгоритмической сложностью решения задач поиска и оптимизации на *конечных множествах*. Классификация подобных задач по сложности стала в последние годы приоритетным направлением. При этом некоторые старые проблемы вновь оказались в центре внимания.

← Современная дискретная математика

Так, более двух тысячелетий назад александрийский математик Эратосфен отсеивал простые числа из натурального ряда методом, который ныне носит его имя. Древние греки занимались числами, не стремясь извлечь из этого какой-либо практической выгоды, а просто из любви к прекрасному. В двадцатом веке, когда теория чисел из чисто теоретической дисциплины превратилась в раздел, имеющий важные приложения, в частности, в криптографии, распознавание в приемлемое время простоты больших по величине натуральных чисел стало практически востребованной задачей. И вот недавно группой индийских математиков (M. Agrawal, N. Kayal, N. Saxena) для решения этой задачи найден эффективный алгоритм, позволяющий с помощью компьютера распознавать простоту чисел, десятичная запись которых составляет сотни цифр.

Традиционный, рассчитанный на механиков курс высшей математики, основу которого составляет дифференциальное и интегральное исчисление, сегодня уже не может в полной мере удовлетворять потребность высшей школы в математических знаниях. Первым учебным пособием, отражающим начавшиеся изменения в преподавании высшей математики, была переведенная на русский язык книга: Дж. Кемени, Дж. Снелл, Дж. Томпсон «Введение в конечную математику». Написанная более полувека назад, она и сейчас сохраняет определённую ценность и может использоваться в учебном процессе. К сожалению, она не оказала в своё время существенного влияния на преподавание математики в нашей стране. Чтобы не перегружать школьников, изучающих теперь элементы математического анализа, из программы средней школы была исключена даже традиционно присутствовавшая в ней элементарная комбинаторика.

Дискретная математика в определённой мере проще непрерывной. Её базовые понятия не требуют абстракции предельного перехода, а многие фундаментальные результаты могут быть наглядно продемонстрированы на элементарных примерах. Однако имевший место на протяжении ряда десятилетий перекос в школьном образовании в сторону непрерывной математики привёл к тому, что изучение дискретной математики вызывает значительную трудность у лиц, чьё математическое образование базируется исключительно на *классическом математическом анализе*.

Цельность математического анализа обеспечивается единым подходом к решению широкого круга задач, основанным на использовании производной и первообразной — понятий, существенно использующих свойства континуума. Принцип непрерывности и основанное на нём понятие предела, красной нитью проходящие через весь математический анализ, позволили создать удивительное по красоте и логической стройности здание, в основании которого лежит замена конечного приращения дифференциалом, а вершиной является теория аналитических функций. Хотя строгое логическое построение континуума и непросто, пространство и время дают достаточную интуитивную основу для его восприятия.

В дискретной математике, занимающейся изучением конечных и счётных множеств, подобного единства достигнуто не было. Она распадается на множество разделов со своими собственными задачами и методами. В задачах дискретной математики аналитические методы уступают место анализу многочисленных возможностей и вариантов. Ярким примером является задача о раскраске произвольной карты четырьмя цветами так, чтобы смежные страны не были раскрашены одним цветом. Доказательство того, что такая раскраска всегда возможна, потребовало рассмотрения столь огромного числа возможных вариантов, что полностью осуществить его оказалось возможным лишь с помощью компьютера.

Несмотря на определённую обособленность, дискретная математика связана практически со всеми остальными математическими дисциплинами. Аналитические методы эффективно используются, особенно в методе производящих функций и задачах, связанных с получением асимптотических оценок. Современная алгебра является мощным средством в дискретной математике, в частности, основным аппаратом теории кодирования. Развитие теории вероятностей на начальном этапе, когда она применялась к теории азартных игр и ограничивалась задачами с конечным множеством элементарных исходов, шло параллельно с развитием перечислительной комбинаторики. При этом обе математические дисциплины взаимно обогащали друг друга. Позднее главенствующую роль в теории вероятностей стали играть аналитические методы и пути развития комбинаторики и теории вероятностей разошлись. Однако во второй половине XX века, когда методы теории вероятностей стали мощным источником неконструктивных доказательств в дискретной математике, связь двух дисциплин снова усилилась.

Тематика общего курса дискретной математики для вузов сейчас находится в стадии формирования. При определенной широте трактовки в него включают элементы высшей алгебры, математической логики и теории чисел. В сложившейся ситуации заслуживает внимания уточнение самой концепции предмета дискретной математики. В настоящей книге представлено ядро сложившегося к настоящему времени вузовского курса дискретной математики. Цель книги — дать представление о дискретной математике в целом, познакомить с её задачами и методами, показать их прикладное значение, а также привить навыки самостоятель-

ной работы. Читателю со школьной базовой математической подготовкой предоставлена возможность познакомиться с современной дискретной математикой.

Основу структурного построения книги составляют главы, разбитые на разделы.

Книга как учебный курс

Вводная глава призвана дать начинающему необходимые элементы общей математической культуры, ввести в круг идей и понятий дискретной математики и подготовить его к их восприятию на более высоком уровне в дальнейших разделах книги. Она знакомит с базовыми понятиями теории множеств, теории вероятностей, алгебры и теории чисел, необходимыми при изучении дискретной математики. Приводимые в этой главе исторические сведения, а также эпизоды биографий творцов науки расширяют общий кругозор читателя и стимулируют его интерес к предмету.

Составляющие основное содержание книги Главы 1–5 написаны суше, здесь больше формул и меньше занимательных историй.

В Главе 1 изучается теория перечисления, являющаяся основой дискретной математики. Перечислительная комбинаторика прививает также те навыки работы с конечными множествами, которые необходимы во всех остальных разделах дискретной математики. Методы перечисления представлены в максимально элементарной форме, с многочисленными примерами, иллюстрирующими их применение.

Глава 2 посвящена булевым функциям — материалу, традиционно включаемому в отечественный курс дискретной математики. Наряду с дизъюнктивной и конъюнктивной нормальными формами рассматриваются вопросы полноты систем булевых функций. Заключительный раздел главы посвящён пороговой логике — проблематике, которая мало освещалась в учебной литературе. Этот раздел может быть использован для спецкурса.

В Главе 3 рассматриваются графы, язык которых ныне широко используется во всех разделах дискретной математики.

Глава 4 посвящена алгоритмам в дискретной математике. Рассматриваются алгоритмы на графах и другие задачи дискретной оптимизации, даётся представление о современном состоянии теории алгоритмической сложности.

Завершающая основную часть Глава 5 посвящена теории кодирования — области дискретной математики, чрезвычайно важной в прикладном отношении и богатой используемыми здесь математическими методами. Рассматриваются задачи сжатия информации, помехоустойчивого кодирования и криптографии.

Более специальные темы вынесены в два «Дополнения», посвящённые общей теории частично упорядоченных множеств и использованию методов теории вероятностей в дискретной математике. Эти дополнения основаны на научных статьях и специальной литературе последних десятилетий и адресованы в первую очередь тем, кто планирует связать с математикой свою будущую профессиональную деятельность.

Вся книга и отдельные её части построены по принципу «от простого — к сложному». Она в равной степени адресована как любознательному старшекласснику, так и работающему над диссертацией аспиранту. Вводная глава является в ней мостиком, ведущим от школьного курса математики к проблемам современной дискретной математики. Начальные разделы глав 1 и 2 могут служить дополнением к школьному курсу математики и информатики, куда теперь включены элементы комбинаторики, теории вероятностей и логики. Для понимания начальных разделов остальных глав книги также в основном достаточно школьного курса математики. Такие понятия математического анализа, как производная, интеграл, ряд, эпизодически появляясь, не играют в общей структуре книги заметной роли. Можно сказать даже, что начало каждой главы написано на гуманитарном уровне. В то же время заключительные разделы глав требуют от читающего уже более серьёзной математической подготовки и соответствуют уровню студента-старшекурсника или аспиранта.

Пять глав основной части в значительной степени независимы друг от друга. Каждая из них, начинаясь совершенно элементарно, затем постепенно переходит к более сложным вещам. Таким образом, знакомство с первыми разделами каждой главы может оказаться достаточным для беглого знакомства с излагаемым в ней предметом.

Для чтения Дополнения 2 требуется владение лишь теми азами теории вероятностей, которые с лихвой покрываются стандартным вузовским курсом. А так как все используемые здесь понятия теории вероятностей вводятся и теоремы доказываются, то этот раздел, в принципе, доступен и тем, кто не слушал (или плохо слушал) такой курс. Первое знакомство читателя с теорией вероятностей происходит во Вводной главе, в разделе 1.4 это знакомство углубляется, а в Дополнении 2 изложение элементов теории случайных величин ведётся уже на стандартном вузовском уровне.

Единственным нарушением положенного в основу книги принципа «от простого — к сложному» является идущий после технически сложного материала заключительный раздел Главы 5, посвящённый криптографии. Он написан в свободной, легко читаемой манере, с многочисленными историческими примерами. Цель подобного изложения состоит в том, чтобы показать криптографию в её историческом развитии и сделать идеи современной криптографии доступными максимально широкому кругу читателей. В качестве предварительной математической подготовки этот раздел предполагает у читателя лишь владение модульной арифметикой.

Помещённые в конце большинства разделов книги «Вопросы для самопроверки» призваны оказать читателю помощь в усвоении материала. Большое число разобранных в тексте примеров, а также задачи для самостоятельного решения, снабжённые ответами и указаниями, дополняют теоретический материал и предоставляют возможность для самостоятельного изучения предмета. Лишь незначительная часть предлагаемых для решения задач придумана самим автором. Большая же часть заимствована

из самых разных источников, не все из которых автор был бы в состоянии теперь указать.

◆ Обозначения и нумерация

В выборе обозначений автор следовал традициям отечественной школы. Нумерация утверждений, лемм, теорем и рисунков начинается заново в каждом разделе каждой главы. При ссылке внутри раздела указывается лишь внутренний номер, при ссылке внутри главы — номер раздела и внутренний номер, а при ссылке из другой главы — номер главы, раздела и внутренний номер.

Книга ориентирована на широкую аудиторию читателей, от любознательных школьников старших классов до студентов, аспирантов и преподавателей вузов, а также всех любителей математики, имея в виду, однако, лиц, интересующихся в математике не только готовыми формулами, но и методами их получения. Поэтому большинство формулируемых в тексте утверждений приведены с полными доказательствами, окончание которых отмечается значком □.

Автор всегда считал полезным как можно раньше знакомить учащихся с открытыми проблемами математики, чтобы они могли почувствовать передний край науки, а наиболее амбициозные из них и попробовать там свои силы. Дискретная математика имеет огромный запас подобного рода нерешённых задач, формулировки которых подчас достаточно элементарны, чтобы позволить начинающему ухватить суть проблемы. Некоторые из них приведены на страницах этой книги.

Как видно из вышесказанного, книга не является жёстким учебным курсом, предполагающим изучение «от корки и до корки». Представленный материал в основном покрывает университетский общий курс дискретной математики, а также может быть основой для аспирантского теорминимума. При составлении же стандартного семестрового курса по дискретной математике, читаемого на втором курсе технического вуза, лектором может быть использовано четыре первых раздела глав 1 и 2, семь первых разделов главы 3 и по четыре первых раздела глав 4 и 5. Материал остальных разделов может включаться в зависимости от уровня математической подготовки аудитории, целей и задач курса. Это же относится и к «Дополнениям». Материал, не включённый в основной курс, может читаться факультативно и использоваться для спецкурсов.

◆ Библиография

В библиографию включены книги, дополняющие и углубляющие изложенный материал, а также те имеющиеся на русском языке работы, в которых впервые появились излагаемые результаты. Знакомство с первоисточниками приоткрывает для учащихся пути развития науки, расширяет их кругозор и прививает вкус к самостоятельному математическому творчеству. Как говорится, «чем ближе к истокам — тем меньше воды». С учётом характера настоящего издания библиография составлена в основном из русскоязычных источников, и в неё включено лишь небольшое число англоязычных книг и статей. Но, разумеется, тем, кто хо-

чет сам активно работать в дискретной математике, необходимо читать научную литературу в оригинале, чтобы быть в курсе последних достижений в этой области.

Благодарности ➡

Автор благодарит помогавших ему в работе над книгой: Н. Н. Кузюрина, с которым он постоянно консультировался в процессе работы, И. П. Чухрова, оказавшего помощь при написании Дополнения 2, В. К. Леонтьева и М. Н. Вялого, сделавших ряд ценных замечаний, а также своих друзей Ю. Г. Свириденко и А. Н. Ходатаева, без помощи которых книга не могла быть написана.

Глава 0

Вводная

Множества.

Множество — это собрание определённых и различимых между собой объектов нашей интуиции или интеллекта, мыслимых как единое целое.

Так определил важнейшее для математики понятие основоположник теории множеств Георг Кантор (1845–1918) во второй половине XIX века. Составляющие множество объекты называются элементами множества. Иногда, руководствуясь геометрическими представлениями, их называют также точками.

Само понятие множества, попыткой прояснить которое является приведенное определение, давно укоренилось в нашем сознании и отражено в естественном языке. Так, мы говорим о компании людей, косяке рыб, стае птиц и т. д. Если основным множеством, с которым имеет дело математический анализ, является множество действительных чисел, то в центре внимания дискретной математики находятся конечные множества, т. е. множества, состоящие из конечного числа элементов.

Общий метод задания произвольного множества состоит в формулировке некоторого характеристического свойства, которым обладают элементы множества и только они. Например, множество натуральных чисел, делящихся на 5, или множество людей, проживающих в определенном населённом пункте. Первое из множеств бесконечно, второе, очевидно, конечно. Конечное множество A может быть, в принципе, задано и простым перечислением своих элементов в произвольном порядке, которое принято записывать в фигурных скобках: $A = \{a_1, \ldots, a_n\}$. Таким образом, $\{a_1, a_2, a_3\}$ и $\{a_3, a_2, a_1\}$ обозначают одно и то же — множество, состоящее из трёх элементов: a_1, a_2, a_3. Если A — множество, состоящее из n элементов, то говорят, что *мощность* множества A равна n и пишут $|A| = n$. Тот факт, что a_i является элементом множества A, записывается как $a_i \in A$.

← Задание множества

В математике рассматривается и множество, не содержащее ни одного элемента. Оно называется пустым множеством и обозначается $\varnothing$.

Так, например, если некоторое уравнение не имеет ни одного решения, то говорят, что множество его решений пусто или, что оно есть пустое множество.

Перестановки.

Список элементов множества, заключённый в круглые скобки, обозначает перечисление элементов в определённом порядке. Такое перечисление в заданном порядке называется перестановкой элементов множества.

Таким образом, (a_1,a_2,a_3) и (a_3,a_2,a_1) обозначают различные перестановки элементов множества $\{a_1,a_2,a_3\}$. Вот все 6 возможных перестановок элементов множества $\{a_1,a_2,a_3\}$:

$$(a_1,a_2,a_3),(a_1,a_3,a_2),(a_2,a_1,a_3),(a_2,a_3,a_2),(a_3,a_1,a_2),(a_3,a_2,a_1).$$

Их число можно было найти и не выписывая в явном виде все 6 перестановок, а с помощью следующего несложного рассуждения. На первое место можно поставить любой из трёх элементов, что даёт 3 возможности. Если на первое место уже поставлен некоторый элемент, то на второе место можно поставить любой из оставшихся двух элементов, что даёт 2 возможности. И, наконец, на третье место остаётся единственный представитель, т. е. единственная возможность. Таким образом, число перестановок оказывается равным $3\times2\times1=6$.

Если бы переставлялись элементы множества мощности n, то число перестановок было бы равно $n\cdot(n-1)\cdot\ldots\cdot2\cdot1$. Для такого произведения натуральных чисел от 1 до n в математике существует специальное обозначение: $1\cdot2\cdot\ldots\cdot(n-1)\cdot n=n!$.

Факториал ➡

Здесь введена важнейшая для комбинаторного анализа функция $n!$ (читается «эн-факториал»), определяемая для натуральных n как произведение всех натуральных чисел от 1 до n включительно и полагаемая равной 1 при $n=0$:

$$0!=1!=1,\ 2!=1\cdot2=2,\ 3!=1\cdot2\cdot3=2!\cdot3=6,\ \ldots,\ n!=(n-1)!\cdot n.$$

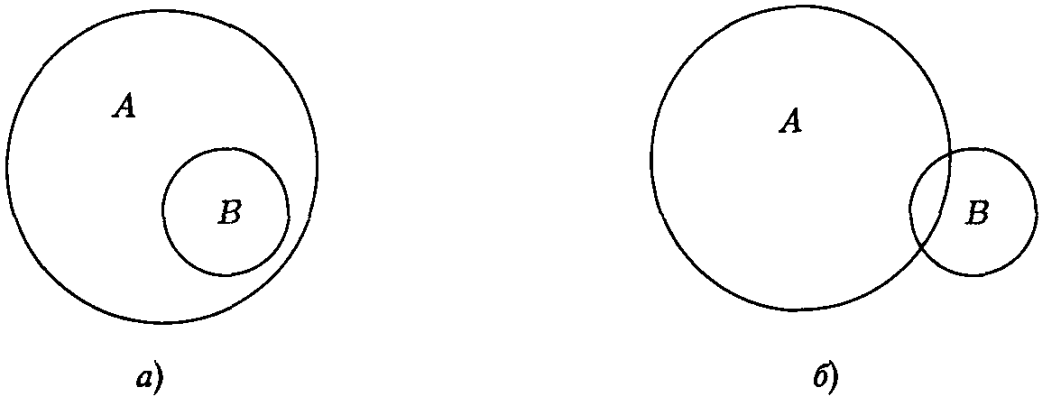
Подмножества.

Если каждый элемент множества B является элементом множества A, то множество B называется подмножеством множества A, что записывается как $B\subseteq A$.

Равенство двух множеств $A=B$ означает, что множества A и B совпадают. Это эквивалентно тому, что $A\subseteq B$ и $B\subseteq A$. Если $B\subseteq A$ и $B\neq A$, то это записывается как $B\subset A$.

Считая элементы множества точками плоскости, лежащими внутри кругов, отношение включения можно наглядно представить как расположение одного круга внутри другого. На рис. 1а $B \subset A$, а на рис. 1б $B \not\subset A$. Такие наглядные представления множеств часто называют диаграммами Эйлера—Венна.

← Диаграммы Эйлера—Венна



а) *б)*

Рис. 1

Подмножествами A являются пустое множество $\varnothing$, не содержащее ни одного элемента, и само множество A. Эти два подмножества называют *несобственными*, а все остальные — *собственными*. Если на множестве A с помощью некоторой перестановки $(a_1,\ldots,a_n)$ зафиксирован порядок элементов, то любое подмножество $B \subseteq A$ можно задать с помощью двоичного набора $(\beta_1, \ldots, \beta_n)$, где $\beta_i = 1$, если $a_i \in B$, и $\beta_i = 0$, если $a_i \notin B$. Такой вектор называется *характеристическим вектором* подмножества B. Например, если $A = \{a_1, a_2, a_3, a_4, a_5\}$, а $B = \{a_1, a_3, a_5\}$, то характеристический вектор подмножества B равен (10101).

← Характеристический вектор

Между подмножествами множества A и двоичными наборами длины $|A|$ существует, таким образом, взаимно однозначное соответствие. Поэтому, пересчитав число двоичных наборов длины $|A|$, можно найти и полное число подмножеств множества A. Так как каждая компонента набора независимо принимает одно из двух значений, то число подмножеств равно $2^{|A|}$. Этим объясняется обозначение 2^A для семейства всех подмножеств множества A, используемое как для конечных, так и для бесконечных множеств.

Таким образом, полное число подмножеств n-элементного множества равно 2^n. Вот все $2^3 = 8$ подмножеств трёхэлементного множества $A = \{a_1, a_2, a_3\}$:

$$2^A = \{\varnothing, \{a_1\}, \{a_2\}, \{a_3\}, \{a_1, a_2\}, \{a_1, a_3\}, \{a_2, a_3\}, \{a_1, a_2, a_3\}\}.$$

Характеристические векторы этих подмножеств есть:

(000), (100), (010), (001), ((110), (101), (011), (111).

Система подмножеств $B_1, B_2, \ldots, B_k \subseteq A$ называется разбиением множества A, если каждый элемент множества A принадлежит в точности одному из подмножеств $B_1, B_2, \ldots, B_k$.

На рис. 2 показано такое разбиение множества на три подмножества.

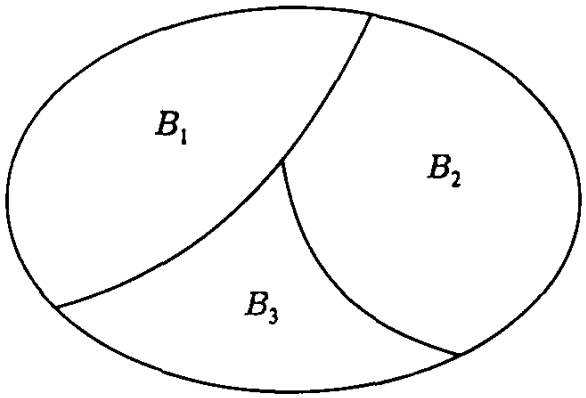


Рис. 2

Счётные множества. Если некоторое множество состоит из n элементов, то его элементы можно занумеровать числами от 1 до n или, как говорят, поставить во взаимно однозначное соответствие с начальным отрезком ряда натуральных чисел $1, 2, \ldots, n$. Если не ограничивать мощности рассматриваемых конечных множеств, то для такой нумерации потенциально необходимо всё множество натуральных чисел $\mathbb{N} = \{1, 2, 3, \ldots\}$.

Множество $\mathbb{N}$ — это бесконечное множество. Когда хотят охарактеризовать его мощность, то говорят, что оно *счётно*. Если между некоторым множеством A и множеством $\mathbb{N}$ можно установить *взаимно однозначное соответствие*, т. е. перенумеровать элементы множества A так, что будут использованы все натуральные числа, то говорят, что множество A равномощно множеству натуральных чисел и также имеет счётную мощность.

Взаимно однозначное соответствие ➡

В общем случае *взаимно однозначным соответствием* $M \leftrightarrow M'$ между двумя множествами M и M' называется такое объединение их элементов в пары $\{m, m'\}$, где $m \in M$, $m' \in M'$, что каждый элемент множества M находится в паре ровно с одним элементом множества M', а каждый элемент множества M' — ровно с одним элементом множества M. Взаимно однозначное соответствие между множествами M и M' можно наглядно представить, соединив линиями элементы двух множеств, объединённые в пары (рис. 3).

Принцип взаимно однозначного соответствия может быть использован при сравнении мощностей конечных множеств. В самом деле, чтобы сравнить число студентов в аудитории с числом имеющихся в ней стульев, можно, не пересчитывая отдельно студентов и стулья, попросить студентов занять свободные стулья. Если все студенты усядутся и свободных стульев

не останется, то установленное таким образом взаимно однозначное соответствие будет означать, что число студентов равно числу стульев. Этот же принцип был положен Кантором и в основу сравнения бесконечных мномножеств по мощности. Согласно Кантору два бесконечных множества *равномощны*, если между ними можно установить взаимно однозначное соответствие.

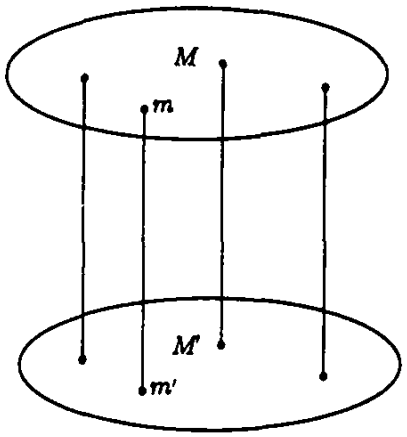


Рис. 3

При таком определении, однако, открывается неожиданное свойство бесконечных множеств, качественно отличающее их от конечных множеств. Бесконечное множество может быть равномощно своему собственному подмножеству. Вот подтверждающий это пример, впервые приведённый в 1638 году великим итальянским мыслителем Галилео Галилеем (1564–1642) в его «Беседах и математических доказательствах, касающихся двух новых отраслей науки, относящихся к механике». Галилей, которому физика обязана тем, что из созданной Аристотелем умозрительной и схоластической науки превратилась в науку экспериментально проверяемую, сделал гениальное, намного опередившее своё время открытие и в математике.

◆ Галилей о бесконечности

Соотношением $i \leftrightarrow i^2$, $i = 1, 2, \ldots$ устанавливается взаимно однозначное соответствие между множеством натуральных чисел и множеством их квадратов, которое, таким образом, также оказывается счётным. Данное взаимно однозначное соответствие может быть наглядно представлено в виде следующей бесконечной таблицы, в верхней строке которой стоят все натуральные числа, а в нижней — их квадраты:

$$\begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & \cdots \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & & . \\ 1 & 4 & 9 & 16 & 25 & 36 & 49 & 64 & \ldots \end{array}$$

Таблица наглядно демонстрирует равномощность двух множеств, хотя второе из них является подмножеством первого. Пример Галилея явился первым шагом в логическом познании бесконечности. Показав этим примером, что у бесконечного часть может быть равна целому, Галилей устами одного из участников беседы резюмирует:

> Свойства равенства, а также большей и меньшей величины не имеют места там, где дело идёт о бесконечности, и применимы только к конечным множествам.

Потребовалось два с половиной столетия на то, чтобы, не испугавшись кажущегося противоречия со здравым смыслом, решиться сравнивать бесконечные множества по мощности. Этот решительный шаг и был сделан Кантором.

Счётность множества A эквивалентна тому, что его элементы можно выписать в бесконечную последовательность, номер элемента в которой равен соответствующему этому элементу натуральному числу. Множество целых чисел $Z = 0, \pm 1, \pm 2, \ldots$ является счётным множеством, так как его элементы можно выписать в бесконечную последовательность, например, следующим образом $0,\ 1, -1,\ 2, -2,\ 3, -3, \ldots$.

Множество рациональных чисел $Q = p/q$, где $p \in Z$, $q \in N$ и p/q — несократимая дробь, также является счётным множеством. Его можно перечислить, последовательно выписывая рациональные числа с $|p| + q = 1$, $|p| + q = 2$, $|p| + q = 3$ и т. д.

Континуум. В любом бесконечном множестве можно выделить счётное подмножество, поэтому можно сказать, что счётное множество — это наименьшее бесконечное множество. Но возникает вопрос, а существуют ли несчётные бесконечные множества, т. е. такие множества, элементы которых невозможно пересчитать, занумеровав натуральными числами? Или же каждое бесконечное множество является счётным? Оказывается, что несчётные бесконечные множества действительно существуют!

Рассмотрим множество всех бесконечных последовательностей, составленных из десятичных цифр, т. е. последовательностей вида $a_1, a_2, a_3, \ldots$, где $a_i \in \{0,1,2,3,4,5,6,7,8,9\}$. Это множество несчётно. Чтобы доказать это, допустим противное. Пусть все такие последовательности можно выписать в виде идущего сверху вниз бесконечного ряда:

$$a_1^1, a_2^1, a_3^1, a_4^1, a_5^1, \ldots$$
$$a_1^2, a_2^2, a_3^2, a_4^2, a_5^2, \ldots$$
$$a_1^3, a_2^3, a_3^3, a_4^3, a_5^3, \ldots$$
$$\ldots\ldots\ldots\ldots\ldots$$

Построим последовательность $b_1, b_2, b_3, \ldots$, руководствуясь следующим правилом: b_1 — любая десятичная цифра, не равная a_1^1; b_2 — любая цифра, не равная a_2^2; b_3 — не равная a_3^3 и т. д. Тогда последовательности $b_1, b_2, b_3, \ldots$ нет в выписанном ряду. Полученное противоречие и доказывает несчётность рассматриваемого множества. Это доказательство было найдено Кантором и называется диагональным методом Кантора. Опираясь на доказанную несчётность множества последовательностей, логично считать, что это множество имеет большую мощность. Мощность рассматриваемого множества последовательностей и всех тех множеств, которые могут быть поставлены во взаимно однозначное соответствие с ним, называется *континуумом*.

Если последовательности $a_1, a_2, a_3, \ldots$ сопоставить бесконечную десятичную дробь $0, a_1 a_2 a_3 \ldots$, то каждой последовательности будет соответствовать некоторая точка отрезка $[0,1]$. Поэтому множество точек отрезка $[0,1]$ также имеет мощность континуум. В переводе с латинского континуум и означает непрерывное. Такую же мощность имеет и множество $\mathbb{R}$ всех действительных чисел. Таким образом, действительных чисел, не представимых в виде p/q, т. е. иррациональных чисел, оказывается существенно больше, чем рациональных. Неудивительно поэтому, что такие естественным образом возникающие в математике действительные числа как $\sqrt{2}$, π, e оказываются иррациональными. Однако доказательство иррациональности в конкретном случае может оказаться сложной задачей. Первым открытым ещё древними греками иррациональным числом было $\sqrt{2}$. С доказательством его иррациональности читатель сможет ознакомиться в конце этой главы.

Иерархия мощностей

Так как $2^n > n$, то число подмножеств любого непустого конечного множества превышает его мощность. В самом деле, число элементов конечного множества A совпадает с числом его одноэлементных подмножеств, т. е. с подмножеством множества 2^A. Это же справедливо и для бесконечных множеств, и мощность 2^A всегда превышает мощность A, однако доказательство этого факта для бесконечных множеств требует более тонкого рассуждения.

Пусть A — произвольное непустое множество, конечное или бесконечное. Допустим противное, пусть между A и 2^A установлено взаимно однозначное соответствие $A \leftrightarrow 2^A$. Определим подмножество $A' \subseteq A$ следующим образом. Для каждого $a \in A$ включим a в A' в том и только в том случае, если a не принадлежит подмножеству, поставленному в соответствие элементу a. Пусть подмножеству A' соответствует элемент a'. Попытаемся ответить на вопрос, принадлежит ли элемент a' подмножеству A'? Если $a' \in A'$, то, по построению множества A', $a' \notin A'$. Если же $a' \notin A'$, то $a' \in A'$, опять-таки по построению множества A'. Таким образом, в любом случае имеет место $a' \in A'$ и $a' \notin A'$. Полученное противоречие и показывает невозможность установления взаимно однозначного соответствия между A и 2^A.

Хотя, как показывает доказанный результат, иерархия мощностей бесконечна, практически все встречающиеся в математике множества конечны, счётны или континуальны. Во времена Кантора математиков интересовали главным образом бесконечные множества, сто лет спустя они проявили интерес и к конечным множествам.

Операции над множествами. Если заданы два множества A и B, то из них по определённым правилам может быть построено новое множество. Это называется теоретико-множественной операцией или просто операцией над множествами.

Объединение и пересечение ➡

Множество, состоящее из элементов, входящих хотя бы в одно из множеств A или B, называется *объединением* множеств A и B и обозначается $A \cup B$.

Множество, состоящее из элементов, входящих в оба множества A и B, называется *пересечением* множеств A и B и обозначается $A \cap B$. Если $A \cap B = \varnothing$, т. е. у множеств A и B нет общих элементов, то говорят, что множества A и B не пересекаются.

Операции объединения и пересечения могут быть наглядно представлены как показано на рис. 4.

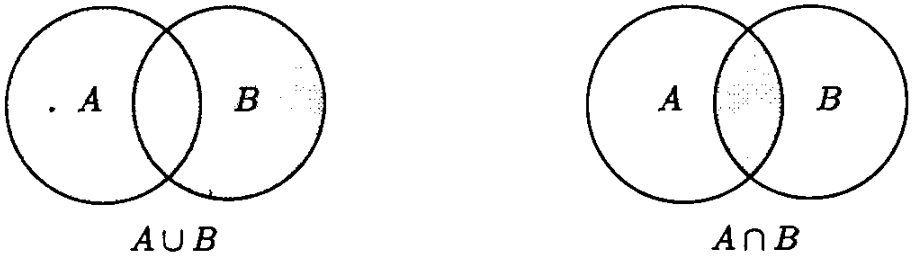


Рис. 4

На практике часто возникает следующая задача. По известным мощностям множеств A и B и мощности их пересечения требуется найти мощность их объединения. Если множества A и B не пересекаются, то $|A \cup B| = |A| + |B|$. Если же пересечение A и B не пусто, то в сумму $|A| + |B|$ каждый элемент пересечения входит дважды. Поэтому в общем случае для подсчёта мощности объединения используется формула

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

В качестве примера использования данной формулы рассмотрим следующую задачу. В группе 20 студентов, каждый из которых обязан изучать хотя бы один из двух иностранных языков — английский или испанский. Известно, что занятия по английскому посещают 15 человек, а по испанскому — 10. Сколько студентов изучают оба иностранных языка?

Ответ на этот вопрос может быть легко получен с помощью вышеприведённой формулы:

$$20 = 15 + 10 - x,$$

откуда $x = 5$.

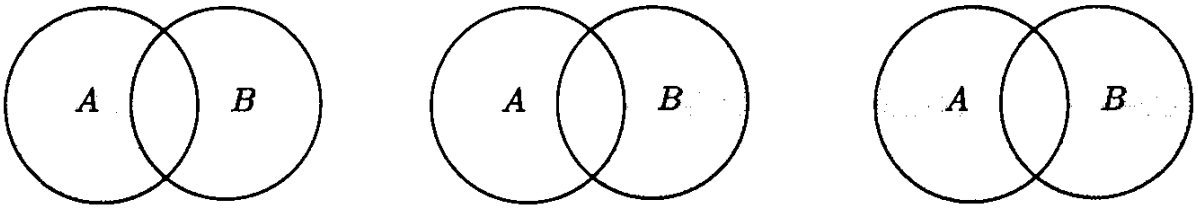
Разность и симметрическая разность ➡

В теории множеств рассматриваются ещё две операции над множествами, которые называются *разность* и *симметрическая разность*.

Множество, состоящее из элементов, входящих в A, но не входящих в B, называется *разностью* множеств A и B и обозначается $A \setminus B$.

Множество, состоящее из элементов, входящих ровно в одно из множеств A или B, называется их *симметрической разностью* и обозначается $A \oplus B$ или $A \Delta B$. Как следует из определения, $A \oplus B = (A \setminus B) \cup (B \setminus A)$, а также $A \oplus B = (A \cup B) \setminus (A \cap B)$. Ясно, что $A \oplus B = B \oplus A$, чем и объясняется название операции.

Разность и симметрическая разность показаны на рис. 5.



$A \setminus B$ $B \setminus A$ $A \oplus B$

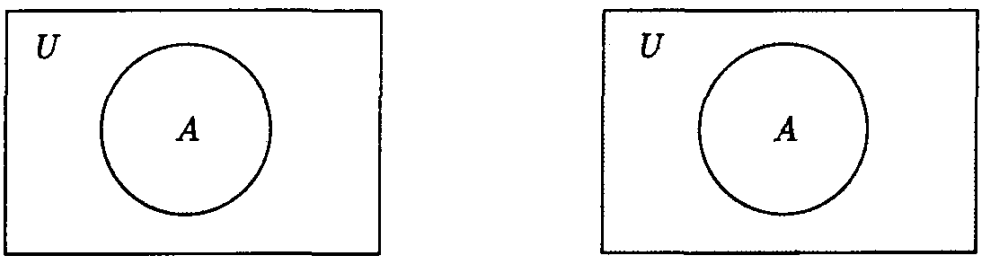
Рис. 5

Как правило, множества, с которыми приходится работать, являются подмножествами некоторого универсального множества. В теории чисел таким множеством является множество натуральных чисел $\mathbb{N} = \{1, 2, 3, \ldots\}$, в математическом анализе — множество действительных чисел $\mathbb{R}$, а в теории аналитических функций — множество комплексных чисел $\mathbb{C}$.

Если обратиться к жизненным примерам, то здесь дело обстоит сходным образом. Так, множество людей, проживающих в определенном населённом пункте, можно рассматривать как подмножество людей, проживающих в государстве, или подмножество вообще всех людей на Земле.

Если такое универсальное для данного круга задач множество U задано, то все рассматриваемые множества будут его подмножествами и для любого из них определена операция *дополнения*: $\bar{A} = U \setminus A$. Схематически она представлена на рис. 6.

◆ Дополнение



A $\bar{A}$

Рис. 6

Разность и симметрическая разность могут быть выражены через основные операции $\cup$, $\cap$, $\bar{}$ как:

$$A \setminus B = A \cap \overline{B}\ ;\ B \setminus A = B \cap \overline{A}\ ;\ A \oplus B = (A \cap \overline{B}) \cup (\overline{A} \cap B).$$

Свойства операций ➡

Операции $\cup$, $\cap$, $\overline{\ }$ обладают рядом интересных свойств. Укажем важнейшие из них.

1. Коммутативность (переместительный закон):

$$A \cup B = B \cup A\ ;\ A \cap B = B \cap A.$$

2. Ассоциативность (сочетательный закон):

$$(A \cup B) \cup C = A \cup (B \cup C)\ ;\ (A \cap B) \cap C = A \cap (B \cap C).$$

3. Дистрибутивность (распределительный закон):

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)\ ;\ A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

4. Инволютивность дополнения (преобразование периода 2, повторное применение которого даёт исходное множество):

$$\overline{\overline{A}} = A.$$

5. Правила де Моргана:

$$\overline{A \cup B} = \overline{A} \cap \overline{B}\ ;\ \overline{A \cap B} = \overline{A} \cup \overline{B}.$$

Свойства 1–3 напоминают свойства арифметических операций «+» и «×», но в отличие от них между операциями $\cup$ и $\cap$ существует полная симметрия. Ассоциативность, коммутативность и инволютивность вполне очевидны и не нуждаются в доказательстве. Справедливость дистрибутивности и правил де Моргана может быть непосредственно усмотрена из диаграмм Эйлера—Венна или установлена несложным логическим рассуждением. В качестве примера такого рассуждения докажем первое из правил де Моргана: $\overline{A \cup B} = \overline{A} \cap \overline{B}$.

Пусть $x \in \overline{A \cup B}$. Тогда

$$x \notin A \cup B \ \Rightarrow\ x \notin A \ и\ x \notin B \ \Rightarrow\ x \in \overline{A}\ и\ x \in \overline{B} \ \Rightarrow\ x \in \overline{A} \cap \overline{B}.$$

Поэтому $\overline{A \cup B} \subseteq \overline{A} \cap \overline{B}$.

Пусть $y \in \overline{A} \cap \overline{B}$. Тогда

$$y \in \overline{A}\ и\ y \in \overline{B} \Rightarrow y \notin A\ и\ y \notin B \Rightarrow y \notin A \cup B \Rightarrow y \in \overline{A \cup B}.$$

Поэтому $\overline{A} \cap \overline{B} \subseteq \overline{A \cup B}$.

Отсюда вытекает, что $\overline{A \cup B} = \overline{A} \cap \overline{B}$.

Прямое произведение. Помимо определённых выше операций, имеется ещё одна важная операция над множествами, которая называется *прямым* или *декартовым произведением множеств*. *Прямым* (*декартовым*) произведением $A \times B$ множеств A и B называется множество всех упорядоченных пар (a,b), где $a \in A$, $b \in B$.

Пусть $A=\{a_1,a_2,a_3\}$, $B=\{b_1,b_2\}$. Тогда их прямое произведение $A\times B=\{(a_1,b_1),(a_1,b_2),(a_2,b_1),(a_2,b_2),(a_3,b_1),(a_3,b_2)\}$ состоит из $3\cdot 2=6$ элементов. И в общем случае мощность прямого произведения двух конечных множеств находится как произведение их мощностей:

$$|A\times B|=|A|\cdot|B|.$$

Множество двоичных векторов длины n с компонентами 0 или 1 или, как говорят, бинарных наборов можно рассматривать, таким образом, как $\{0,1\}^n$ — n-ю декартову степень множества $\{0,1\}$. Поэтому число таких наборов в соответствии с вышеприведённой формулой равно 2^n, что уже было ранее доказано.

✓*Декарт*

Само название «декартово произведение» происходит от имени великого французского учёного и философа Рене Декарта (1596–1650), который в историю математики вошёл как создатель аналитической геометрии. Если геометрическим образом множества действительных чисел $\mathbb{R}$ является прямая, то геометрический образ декартова квадрата $R\times R=R^2$ является плоскость, так как каждая точка плоскости задаётся в аналитической геометрии упорядоченной парой действительных чисел (x,y) — абсциссой и ординатой точки.

В 1637 году Декарт издал свой основной философский труд, знаменитое «Рассуждение о методе, чтобы хорошо направлять свой разум и отыскивать истину в науках», к которому прибавил три приложения: «Диоптрика», «Метеоры» и «Геометрия», призванные практически разъяснять его «метод». Именно в его «Геометрии» были заложены основы аналитической геометрии на плоскости, связавшей алгебру с геометрией. Благодаря открытию Декарта геометрические линии, в том числе и такие хорошо известные и изученные ещё античными геометрами как прямая, окружность, эллипс, гипербола, парабола, могли задаваться алгебраическими уравнениями и изучаться алгебраическими методами. Открытие Декарта до сих пор, в виде уже алгебраической геометрии, является мощным стимулом для математических исследований.

Начало его жизни, казалось, не предвещало ничего необычного. После окончания находившегося под покровительством Генриха IV иезуитского коллежа в Ла-Флеше он отправился в Париж, где повёл типичную светскую жизнь молодого повесы с кутежами, попойками и азартными играми. Но короткое время спустя резко изменил образ жизни, уединился и засел за изучение наук. Видя это, родные, не терявшие надежды образумить молодого человека, сделали попытку его женить. Хорошенькая барышня, которую ему прочили в жёны, завязала с Декартом разговор о «различных видах красоты». Но Декарт вместо ожидаемого комплимента заметил, что «из всех известных ему видов красоты на него наиболее сильное впечатление произвела красота Истины».

Декарт сильнейшим образом повлиял на зарождавшуюся европейскую науку и философию. Ему принадлежит знаменитый тезис:

Cogito, ergo sum (мыслю, значит существую).

Вероятность. Элементарная теория вероятностей тесно связана с мощностями конечных множеств, и здесь уместно коснуться её основных принципов.

Когда имеют дело с опытом, результат которого заранее предсказать невозможно, но говорят о вероятности того или иного исхода. Если бросается симметричная монета, то вероятность выпадения орла равна $1/2$, так как в данном опыте всего возможны два исхода, а нас интересует один из них. Вероятность $1/2$ означает, что при многократном подбрасывании монеты доля выпадений орла составит примерно половину от общего числа подбрасываний.

Пусть теперь бросается игральная кость, и нас интересует вероятность события, что выпадет не менее 5 очков. При бросании кости, имеющей шесть граней, возможны 6 элементарных исходов: 1, 2, 3, 4, 5, 6, а нашему событию благоприятствуют два из них: 5 и 6. Поэтому его вероятность равна $2/6 = 1/3$, т. е., в среднем, на каждые три броска один раз происходит интересующее нас событие. Аналогичным образом могут быть вычислены вероятности и других событий в азартных играх — как отношение числа элементарных исходов, благоприятствующих событию, к полному числу возможных элементарных исходов данного опыта.

Вероятность по Лапласу ➡

Именно так определил это понятие в своём фундаментальном труде «Аналитическая теория вероятностей» в 1812 году Пьер Симон Лаплас (1749–1827) — выдающийся французский математик, физик, астроном и философ, внёсший значительный вклад во многие разделы математики, математической физики и небесной механики.

Согласно Лапласу, вероятность $\mathrm{P}(A)$ события A определяется как

$$\mathrm{P}(A) = \frac{\textit{число элементарных исходов, благоприятствующих } A}{\textit{полное число возможных элементарных исходов}}.$$

Это определение часто называют «классическим определением вероятности». Им можно пользоваться в тех случаях, когда множество элементарных исходов опыта конечно, а соображения симметрии позволяют считать эти исходы равновозможными. Азартные игры в полной мере отвечают этим требованиям. Кроме того, один и тот же опыт здесь многократно воспроизводится при одинаковых условиях, что позволяет извлечь из вычисленной вероятности и практическую пользу.

Классическое определение вероятности может быть изящно выражено на языке теории множеств. Обозначим множество возможных элементар-

ных исходов через Ω, а множество элементарных исходов, составляющих интересующее нас событие, через A, и будем отождествлять событие с этим множеством. Тогда *вероятность* $\mathrm{P}(A)$ *события* A выразится как

$$\mathrm{P}(A) = \frac{|A|}{|\Omega|}.$$

Так как $A \subseteq \Omega$, определённая таким образом вероятность заключена в пределах

$$0 \le \mathrm{P}(A) \le 1.$$

При этом, если $\mathrm{P}(A) = 0$, то событие A рассматривается как *невозможное*, а если $\mathrm{P}(A) = 1$, то как *достоверное*. В нашем определении невозможным событием является $\varnothing$ — пустое множество, а достоверным — всё множество Ω. При промежуточных значениях $\mathrm{P}(A)$ может рассматриваться в качестве меры достоверности события A. При этом уже отмечавшееся свойство вероятности $\mathrm{P}(A)$ состоит в том, что, если один и тот же опыт независимо многократно повторяется при неизменных условиях, то относительная частота осуществления события A (отношение числа осуществлений события A к числу повторений опыта) с ростом n приближается к $\mathrm{P}(A)$. Это важнейшее свойство вероятности называется *законом больших чисел*. Вероятности, таким образом, можно воспринимать как некоторые «идеальные частоты», к которым реальные частоты, флуктуируя, сходятся в процессе многократного повторения опыта.

С каждым событием A связано *дополнительное* к нему или *противоположное событие* $\bar{A} = \Omega \setminus A$, состоящее из всех тех и только тех исходов, которые не входят в A. Имеем

$$\mathrm{P}(\bar{A}) = \frac{|\bar{A}|}{|\Omega|} = \frac{|\Omega \setminus A|}{|\Omega|} = \frac{|\Omega| - |A|}{|\Omega|} = 1 - \frac{|A|}{|\Omega|} = 1 - \mathrm{P}(A).$$

В нашем случае $\Omega = \{1,2,3,4,5,6\}$, $A = \{5,6\}$, $\bar{A} = \{1,2,3,4\}$ — событие, состоящее в том, что выпадет меньше 5 очков, $\mathrm{P}(\bar{A}) = 2/3$.

Но в игре в кости бросается не одна, а несколько костей (обычно две или три). Каково число возможных элементарных исходов, если бросаются две кости? Теперь элементарный исход является упорядоченной парой двух чисел — числа очков, выпавших на первой кости, и числа очков, выпавших на второй кости. Например, элементарный исход $(2,3)$ означает, что на первой кости выпало 2 очка, а на второй — 3. Множеством элементарных исходов в случае бросания двух костей является, таким образом, множество $\Omega \times \Omega$ —декартов квадрат множества элементарных исходов при бросании одной кости, и его мощность равна $6^2 = 36$.

Если мы хотим найти вероятность $\mathrm{P}(A)$ того, что при бросании двух игральных костей выпадет не более 4 очков, то нужно принять во внимание,

что множество элементарных исходов состоит теперь из $6^2 = 36$ исходов, а нашему событию благоприятствуют шесть из них: $(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)$. Поэтому искомая вероятность равна

$$\mathrm{P}(A) = \frac{6}{36} = \frac{1}{6}.$$

Пусть теперь снова бросаются две кости, и нас интересует вероятность одновременного осуществления некоторого события A на первой кости и события B на второй. Пусть, например, событие A состоит в том, что на первой кости выпадет не менее 5 очков, $A = \{5,6\}$, а событие B — что на второй выпадет чётное число очков, $B = \{2,4,6\}$. Тогда событие $A \cap B$, состоящее в одновременном наступлении события A и события B, есть

$$A \cap B = \{(5,2),(5,4),(5,6),(6,2),(6,4),(6,6)\} = A \times B,$$

а его вероятность $\mathrm{P}(A \cap B) = \frac{6}{36} = \frac{1}{6}$. Однако, эта вероятность может быть подсчитана и иным, более удобным способом:

$$\mathrm{P}(A \cap B) = \frac{|A \times B|}{|\Omega \times \Omega|} = \frac{|A| \cdot |B|}{|\Omega| \cdot |\Omega|} = \frac{|A|}{|\Omega|} \cdot \frac{|B|}{|\Omega|} = \mathrm{P}(A) \cdot \mathrm{P}(B) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}.$$

В общем случае, если проводится два независимых опыта с элементарными исходами Ω_1 и Ω_2, и нас интересует вероятность события $A \cap B$, состоящего в одновременном появлении события A в первом опыте и события B во втором, то она также находится как

$$\mathrm{P}(A \cap B) = \frac{|A \times B|}{|\Omega_1 \times \Omega_2|} = \frac{|A| \cdot |B|}{|\Omega_1| \cdot |\Omega_2|} = \frac{|A|}{|\Omega_1|} \cdot \frac{|B|}{|\Omega_2|} = \mathrm{P}(A) \cdot \mathrm{P}(B).$$

Мы получили результат, имеющий важнейшее значение для теории вероятностей:

> Вероятность одновременного осуществления нескольких независимых событий равна произведению их вероятностей.

Лаплас оказал огромное влияние на развитие теории вероятностей, хотя по своим философским взглядам и стоял на позициях крайнего детерминизма, считая саму концепцию вероятности следствием неполноты в наших знаниях причин и условий явлений:

> Все явления, даже те, которые по своей незначительности как будто не зависят от великих законов природы, суть следствия столь же неизбежные этих законов, как обращение солнца. ... Всякое имеющее место явление связано с предшествующим на основании того очевидного принципа, что какое-либо явление не может возникнуть без производящей его причины. ... Таким образом, мы должны рассматривать настоящее состояние вселенной как следствие её предыдущего состояния и как причину последующего.
>
> *Лаплас «Опыт философии теории вероятностей» (1814)*

А вот высказывание Лапласа, ставшее каноническим выражением философского детерминизма:

> Ум, которому были бы известны для какого-либо данного момента все силы, одушевляющие природу, и относительное положение всех её составных частей, если бы вдобавок он оказался достаточно обширным, чтобы подчинить эти данные анализу, обнял бы в одной форме движение величайших тел вселенной наравне с движением легчайших атомов: не осталось бы ничего, что было бы для него недостоверно, и будущее, так же как и прошедшее, предстало бы перед его взором.
>
> *Лаплас «Опыт философии теории вероятностей» (1814)*

За прошедшее с тех пор время научные воззрения существенно изменились, детерминизм отступил, а концепция вероятности, напротив, усилила свои позиции, и согласно современной квантовой физике фундаментальные законы микромира носят существенно вероятностный характер.

Будучи последовательным материалистом и атеистом, Лаплас стал автором первой серьёзной научной космогонической гипотезы, которую представил в своей книге «Изложение системы мира». Говорят, что когда Наполеон, которому Лаплас преподнёс свой труд, спросил его о месте бога в его системе, Лаплас ответил, что он не нуждается в этой гипотезе.

Обладая огромной энергией и честолюбием, Лаплас с приходом в 1799 году к власти Наполеона, которому он пятнадцатью годами ранее читал лекции по математике в артиллерийском училище, занял даже пост министра внутренних дел. Однако на этом поприще он не снискал себе лавров. Впоследствии, находясь уже в изгнании на острове Святой Елены, Наполеон сказал о Лапласе:

> Первоклассный геометр вскоре заявил себя администратором более чем посредственным... он весь был проникнут духом «бесконечно малых», который он вносил и в администрацию.

Зарождение теории вероятностей

Хотя Лапласу и принадлежит данное выше классическое определение вероятности, сама концепция математической вероятности сложилась задолго до него и может считаться одним из выдающихся достижений эпохи Возрождения. Мощный духовный взлёт Возрождения оставил свой след и в математике. Были получены формулы для решения уравнений третьей и четвёртой степени, которые не были известны в эпоху античности, и которые не удалось найти арабским математикам. Но, пожалуй, самым ярким его проявлением стало зарождение новой математической дисциплины, аналога которой не было в античной математике — теории вероятностей.

Именно с азартными играми, получившими особенно широкое распространение в Европе после крестовых походов, связано зарождение теории вероятностей в эпоху Возрождения. Известная с глубокой древности, игра в кости в это время приняла характер пандемии. Два или три кубика выбрасывались на стол, и ставку брал выбросивший большую сумму.

Повального увлечения игрой не смогли остановить даже запреты церковных соборов. В качестве художественного образа её использовал Данте в «Божественной комедии»:

Когда кончается игра в три кости,
То проигравший снова их берёт
И мечет их один, в унылой злости;

Другого провожает весь народ;
Кто спереди зайдёт, кто сзади тронет,
Кто сбоку за себя словцо ввернёт.

А тот идёт и только ухо клонит;
Подаст кому, — идти уже вольней,
И так он понемногу всех разгонит.

Таков был я в густой толпе теней,
Чьё множество казалось превелико,
И, обещая, управлялся с ней.

Данте Алигьери «Божественная комедия» (1307–1321)

Пытливый ум человека эпохи Возрождения не мог смириться с невозможностью проникновения в тайну случая. И вот, в комментариях к венецианскому изданию 1477 года в примечании к этому месту уже появляются подсчёты количества различных исходов при бросании трёх костей, оказавшиеся, правда, ошибочными.

✓*Кардано*

Первым серьёзным математическим исследованием проблемы стала «Книга об игре в кости» (1526) Джироламо Кардано (1501–1576), где знаменитый итальянский математик, механик и философ впервые привёл правильные расчёты шансов. Кардано несомненно был одним из титанов Ренессанса, хотя и с элементами экстравагантности. Он первым ввёл в употребление комплексные числа, его имя носит формула для решения кубического уравнения в алгебре, а также карданный вал и карданов подвес в механике. Справедливости ради следует, правда, отметить, что формула для решения уравнения была получена не Кардано, но была им впервые опубликована в книге «Великое искусство» в 1545 году.

Кардано написаны сотни книг, последней из которых стала его автобиография «О моей жизни», законченная незадолго до смерти в 1576 году. Цитата из неё и была использована в качестве эпиграфа. Не все из написанных им книг были изданы при жизни, и «Книга об игре в кости» увидела свет лишь в 1663 году.

Хотя сейчас Кардано известен нам как математик, механик и криптограф, среди своих современников он прославился как врач и астролог. Среди сотен составленных им гороскопов был и гороскоп Иисуса Христа, что привело к определённым осложнениям в его отношениях с инквизицией. Им был также составлен и собственный гороскоп. Широко распространена и необычайно живуча легенда о том, что, вычислив с его помощью дату

своей смерти, он, чтобы не нарушать Мирового порядка, в назначенный срок лишил себя жизни.

Занимаясь столь разнообразными вещами, Кардано, однако, не считал себя разбрасывающимся, полагая, что «слишком разбросанный ум к постиженью вещей неспособен». «Книга об игре в кости» писалась им в молодом возрасте, когда он частенько пополнял свой скромный студенческий бюджет успешной игрой, за что корил себя потом на склоне лет. Однако именно в этой книге был сделан первый шаг в создании теории вероятностей.

✓де Мере, Паскаль и Ферма

В семнадцатом веке эпидемия азартных игр перекинулась из Италии во Францию. Шевалье де Мере (1607–1648), светский человек, философ и литератор, интересовавшийся также математикой и азартными играми, сообщил Паскалю о своём решении двух задач, связанных с бросанием костей. В первой задаче требовалось найти вероятность выпадения хотя бы одной шестёрки при бросании четырёх костей, а во второй — вероятность выпадения двух шестёрок при 24-кратном бросании двух костей, причём решение второй задачи оказалось у де Мере ошибочным.

Блез Паскаль (1623–1662), великий французский математик и физик, сконструировавший также одно из первых механических вычислительных устройств, сообщил о задачах де Мере другому великому французскому учёному. Юрист из Тулузы Пьер Ферма (1601–1666), занимаясь в свободное от основной работы время наукой, сделал ряд крупнейших открытий в математике и физике. Его Великую теорему в теории чисел удалось доказать лишь в конце XX века. В завязавшейся между Паскалем и Ферма в 1654 году научной переписке произошло дальнейшее становление новой теории.

Мушкетёры ⬅ играют в кости

А теперь обратимся к роману Дюма «Три мушкетёра», герои которого жили во Франции в ту же эпоху и, разумеется, тоже играли в кости. Откроем роман на той странице, где Атос сообщает д'Артаньяну, что проиграл в кости англичанину свою и его лошадь, и предлагает ему отыграться, поставив два седла против одной из лошадей. Далее предоставим слово самих героям романа.

— Послушайте, вы, кажется, давно не играли, д'Артаньян?

— И не имею ни малейшей охоты играть.

— Не зарекайтесь. Итак, говорю я, вы давно не играли, и, следовательно, вам должно везти.

.....

Д'Артаньян, дрожа, бросил кости — выпало три очка; его бледность испугала Атоса, и он ограничился тем, что сказал:

— Неважный ход, приятель. Вы, сударь, получите лошадей с полной сбруей.

Торжествующий англичанин даже не потрудился смешать кости; его уверенность в победе была так велика, что он бросил их на стол не глядя. Д'Артаньян отвернулся, чтобы скрыть досаду.

— Вот так штука, — как всегда спокойно проговорил Атос, — какой необыкновенный ход, я видел его всего четыре раза за всю мою жизнь: два очка!

Проанализируем данный эпизод с помощью теории вероятностей. Два очка выбрасываются при единственном элементарном исходе $(1,1)$, так что вероятность этого события равна $1/36$. Трём очкам соответствуют два элементарных исхода: $(1,2)$ и $(2,1)$. Поэтому вероятность выбросить не более трёх очков равна $3/36 = 1/12$. Атос прав, когда называет бросок д'Артаньяна неважным. После того, как д'Артаньян выбросил три очка, вероятность англичанину проиграть лошадь была равна всего лишь $1/36$. Поэтому его бросок можно действительно назвать катастрофическим.

Кажется, однако, странным, что Атос, регулярно игравший в кости, наблюдал выпадение двух очков всего лишь четыре раза в своей жизни. Два очка выпадает, в среднем, один раз на каждые 36 бросаний, т. е. это событие не является столь уж редким. Разумеется, благородный Атос не мог солгать. Дело здесь, по-видимому, в другом. Во времена Александра Дюма игра в кости уже вышла из моды, и великий романист не мог иметь в ней достаточно опыта, а с теорией вероятностей, не слишком широко известной в его время наукой, он знаком не был.

Добавим к этому, что Шарль Ожье де Батц де Кастельмор, д'Артаньян (1611–1673) — капитан-лейтенант первой роты королевских мушкетёров, удивительные события жизни которого использовал Дюма в своём романе, также избегал играть в кости, не слишком-то доверяя капризам Фортуны. Но однажды, спасая любимую женщину, он решительно поставил на кон весь наличный капитал и ... выиграл, как и в романе![1]

Задачи де Мере ➡

Рассмотрим теперь две задачи де Мере. Вероятность, что 6 не выпадет ни разу при бросании четырёх костей, равна $(5/6)^4 = 625/1296$, а вероятность, что выпадет хотя бы раз: $1 - 625/1296 = 671/1296$, т. е. вероятность выпадения превышает вероятность невыпадения, и в более выгодном положении оказывается игрок, делающий ставку на выпадение шестёрки. Вероятность же непоявления ни разу двух шестёрок при 24 бросаниях равна $(35/36)^{24} \approx 0,509$, а вероятность, что эта комбинация появится хотя бы один раз: $1-(35/36)^{24} \approx 0,491$. Де Мере же ошибочно полагал, что вероятность хотя бы одного выпадения двух шестёрок при 24 бросаниях превышает вероятность невыпадения этой комбинации ни разу.

Геометрический подход ➡

Определив события в теории вероятностей как подмножества универсального множества Ω, можно, опираясь на диаграммы Эйлера—Венна, развить иной, геометрический подход к определению понятия вероятности.

---

[1] Мемуары мессира д'Артаньяна. М., Антанта Лтд, 1995.

Будем считать, что на изображённое на рис. 7 множество Ω бросается точка (подобно дротикам в дартсе, с тем, однако, условием, что дротик всегда попадает в область Ω). Брошенная точка может попасть в любую часть множества Ω, так что все точки Ω равноправны, и попадание в любую заданную область определяется лишь площадью этой области. Считая событием A попадание точки в соответствующую этому событию область, имеем

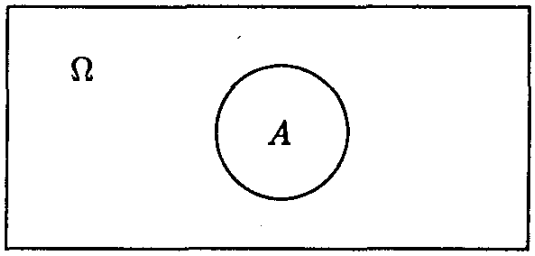


Рис. 7

$$\mathrm{P}(A) = \frac{S_A}{S_\Omega},$$

где S_A и S_Ω — площади областей A и Ω.

Условные вероятности

Геометрический подход позволяет наглядно определить важное понятие *условной вероятности*. Пусть известно, что произошло событие B. Какова теперь вероятность события A (рис. 8)?

Вероятность $\mathrm{P}(A \mid B)$ события A при условии осуществления события B определится теперь как

$$\mathrm{P}(A|B) = \frac{S_{A\cap B}}{S_B} = \frac{S_{A\cap B}/S_\Omega}{S_B/S_\Omega} = \frac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)}.$$

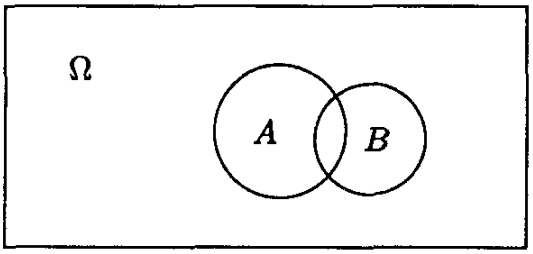


Рис. 8

Используя определение условной вероятности, вероятность одновременного наступления событий A и B можно представить как

$$\mathrm{P}(A\cap B) = \mathrm{P}(B)\cdot\mathrm{P}(A \mid B).$$

Если события A и B независимы, то $\mathrm{P}(A \mid B) = \mathrm{P}(A)$, и формула для вероятности одновременного осуществления событий остаётся прежней

$$\mathrm{P}(A\cap B) = \mathrm{P}(B)\cdot\mathrm{P}(A).$$

Современная теория вероятностей, аксиоматическое построение которой в основном было завершено в 1933 году выдающимся советским математиком академиком А. Н. Колмогоровым (1903–1987), не вдаётся в физический смысл зависимости и независимости, а просто считает два события независимыми, если вероятность их одновременного осуществления равна произведению вероятностей осуществления каждого из них в отдельности. Так, если бросается игральная кость и рассматриваются событие A, состоящее в том, что на ней выпадет не менее 5 очков, и событие B, что выпадет чётное число очков, то эти события оказываются независимыми, так как $\mathrm{P}(A) = 1/3$, $\mathrm{P}(B) = 1/2$ и $\mathrm{P}(A\cap B) = 1/6$. Но если событие A состоит в том, что выпадет не менее 4 очков, то события

A и B оказываются уже зависимыми, так как $P(A) = 1/2$, $P(B) = 1/2$, а $P(A \cap B) = 1/3$.

В жизни также встречаются пары как независимых, так и зависимых событий. Однако здесь при решении вопроса о зависимости или независимости руководствуются, как правило, причинно-следственными связями. Например, зимой существует вероятность образования гололёда, и всегда существует вероятность упасть и сломать ногу. Но в гололёд эта вероятность увеличивается, и данные события нельзя считать независимыми. В то же время образование гололёда и провал на экзамене по математике естественно считать независимыми событиями.

Вернёмся, однако, к геометрическому подходу к вероятности. Вероятности, определяемые такими геометрическими мерами как длина, площадь и объём, называют *геометрическими вероятностями*. Если классическое определение вероятности применимо в задачах с конечным числом равновозможных элементарных исходов, то геометрический подход позволяет рассматривать и случаи с бесконечным числом равноправных элементарных исходов. Если на стол бросается игральная кость, то число элементарных исходов равно шести, но если на плоскость падает шар, то он может коснуться её любой из бесконечного множества точек своей поверхности.

Геометрические вероятности ➡

Пусть теперь некоторая область поверхности шара замелована, и рассматривается событие, состоящее в том, что падение шара оставит на плоскости белую метку. Вероятность этого события определится как отношение площади замелованной поверхности к полной поверхности шара.

Рассмотрим простейшее применение геометрического подхода на примере следующей задачи. Пусть автобусы ходят по маршруту регулярно с интервалом в один час. Некто, не зная расписания, приходит на остановку в случайный момент времени. Какова вероятность, что время ожидания им автобуса будет лежать в интервале от 20 до 40 минут?

Решение этой задачи ясно видно из рис. 9. Время ожидания равномерно распределено на отрезке от 0 до 60. Интересующее нас событие выражается отрезком от 20 до 40, длина которого составляет третью часть исходного. Поэтому и искомая вероятность равна одной третьей.

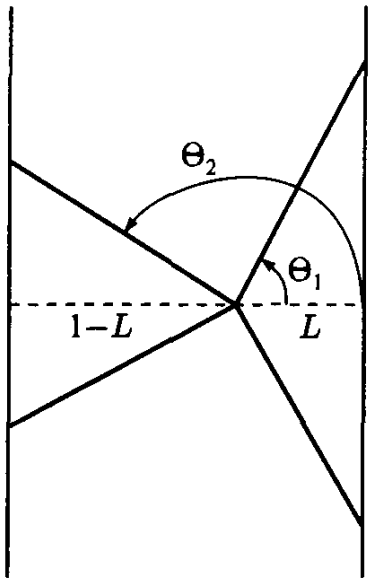
Рис. 9

Самой знаменитой задачей на геометрические вероятности, безусловно, является задача французского натуралиста и естествоиспытателя, директора Королевского ботанического сада в Париже Жоржа Луи Леклерка де Бюффона (1707–1788). Основной труд его жизни — многотомная «Естественная история» был переведён почти на все европейские языки.

Задача Бюффона ➡

Им восхищались А. С. Пушкин и Л. Н. Толстой. Бюффон, однако, интересовался не только растениями и животными. Его имя осталось в теории вероятностей благодаря придуманной им следующей задаче об игле (1777). Игла единичной длины бросается на плоскость, на которой нанесены параллельные прямые, отстоящие друг от друга также на расстоянии единица. Попав в вертикальном положении остриём в случайную точку плоскости, игла затем случайным образом ложится горизонтально. Требуется найти вероятность того, что в этом положении игла пересечёт одну из прямых.

Будем считать параллельные прямые вертикальными. Расстояние от точки падения до ближайшей прямой справа обозначим через L, $0 \le L \le 1$. Тогда расстояние до ближайшей прямой слева будет $1-L$. Будем измерять случайный угол Θ падения иглы на плоскости в обе стороны от направления к правой прямой, так что $-\pi \le \Theta \le \pi$. Как видно из рис. 10, игла пересечёт одну из двух ближайших к точке падения прямых, если угол Θ удовлетворяет одному из двух соотношений $\cos\Theta \le L$ или $\cos\Theta \le L-1$.



$\cos\Theta_1 = L$

$\cos\Theta_2 = L - 1$

Рис. 10

Считая случайные величины L и Θ независимыми и равномерно распределёнными в интервалах своих изменений, получаем, что множество Ω является прямоугольником с основанием 2π и высотой 1, затемнённая область которого определяет интересующее нас событие (рис. 11).

Так как площадь одной арки косинусоиды равна 2, то площадь выделенной области равна 4, и искомая вероятность есть $2/\pi$.

Если теперь многократно бросать иглу на расчерченную плоскость, то в силу закона больших чисел число опытов, закончившихся пересечением, отнесённое к полному числу опытов, будет приближаться к $2/\pi$. В принципе, результаты таких опытов позволяют экспериментально находить число π, хотя для надёжного получения значения $\pi \approx 3{,}14$ требуются тысячи бросаний. В опытах Бюффона можно увидеть в зародыше современные методы

статистических испытаний, объединённые под общим названием методов Монте-Карло. В них роль иглы выполняют уже датчики случайных чисел.

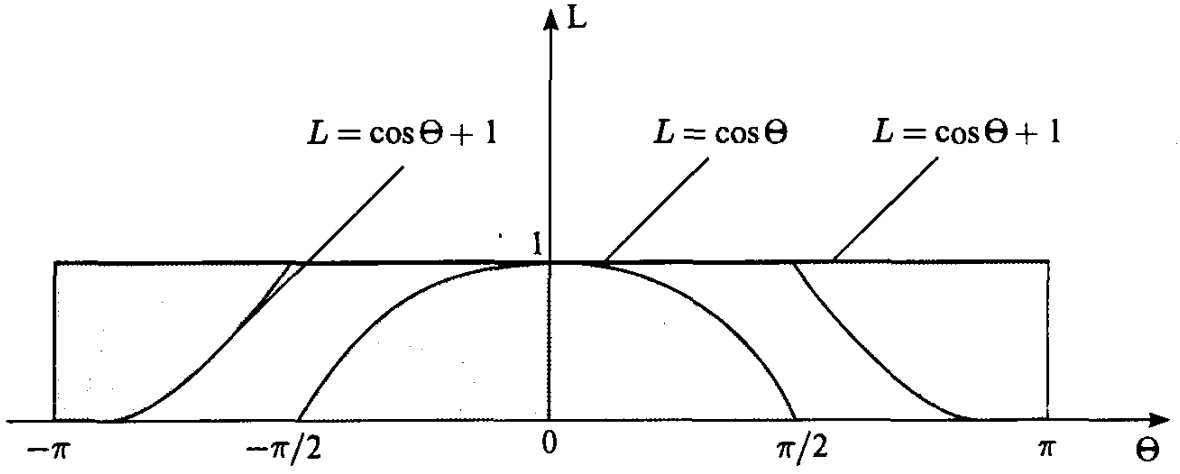


Рис. 11

Теория чисел. Множество натуральных чисел с определёнными на нём операциями сложения и умножения представляет собой важнейшую дискретную структуру в математике и вообще в истории человечества. Греки первыми начали систематическое исследование натурального ряда. В школе Пифагора (VI век до н. э.) число воспринималось как основа мироздания. «Всё есть число» — учили пифагорейцы, и можно сказать, что этот базовый тезис их школы получил своё дальнейшее развитие в современной математической физике. Однако в отличие от нашего времени пифагорейцы связывали с числами все стороны жизни. Так, к примеру, чётные числа олицетворяли женское начало, а нечётные — мужское. Число же 5, равное сумме первого женского и первого мужского числа, было символом брака.

Пифагорейцы ➡

Подтверждение тому, что мир управляется числами, пифагорейцы видели и в совпадающих с первыми членами натурального ряда отношениях длин струны, дающих музыкальные консонансы. Так отношения $1:2$, $2:3$, $3:4$ дают соответственно октаву, квинту и кварту, в чём можно непосредственно убедиться, измерив расстояния между соответствующими ладами на грифе гитары.

Совершенные и дружественные числа ➡

Среди натуральных чисел особым почётом пользовались числа, равные сумме своих собственных делителей, которые получили название *совершенных*. Первыми совершенными числами являются $6 = 1+2+3$ и $28 = 1+2+4+7+14$. Греки предложили и общий способ нахождения таких чисел, который будет рассмотрен в конце этого раздела.

Другой важной, по мнению древних греков, достопримечательностью натурального ряда является наличие в нём пар *дружественных чисел*, каждое из которых равно сумме делителей другого, как, например, $220 = 1+$

$+2+4+71+142$ и $284=1+2+4+5+10+11+20+22+44+55+110$. Говорят, когда однажды Пифагора спросили, что такое друг, он сказал, что это «второе я» и сравнил двух друзей с числами 220 и 284.

Фигурные ⬅ числа

Важную роль в представлении древних греков о числах играло связывавшее их с геометрией понятие *фигурного числа*. Числа-камешки раскладывались в виде правильных геометрических фигур. Среди фигурных чисел выделялись *квадратные*, вида n^2, *треугольные*, вида $1+2+\ldots+n=$ $=n(n+1)/2$ (см. рис. 12), а также ряд других.

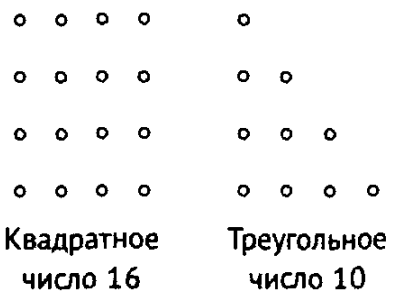


Рис. 12

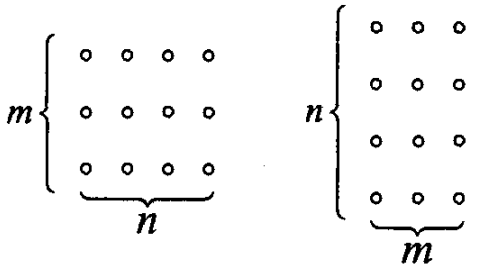
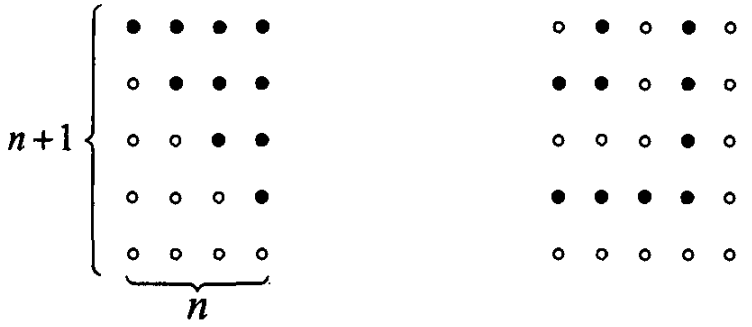


Рис. 13

Фигурное представление чисел позволяет наглядно демонстрировать свойства числового ряда. Так, коммутативный закон умножения $mn=nm$ можно увидеть из двух способов подсчёта камешков в прямоугольнике (рис. 13).

А вот как наглядно с помощью фигурных чисел могут быть просуммированы первые n чисел натурального ряда и первые n нечётных чисел (рис. 14).



$$1+2+3+\ldots+n=\frac{n(n+1)}{2} \qquad 1+3+5+\ldots+(2n-1)=n^2$$

Рис. 14

Изучая числа и фигуры, пифагорейцы считали, что это изучение не должно преследовать каких-либо корыстных целей. Вот сохранённый временем их девиз:

Фигура и шаг вперёд, но не фигура и три обола[1].

[1] Обол — монета в Древней Греции.

Если для пифагорейцев учение о числах было составной частью их религии, включавшей также и веру в переселение душ, то в «Началах» Евклида можно найти уже вполне современный рациональный подход к теории чисел. Ряд основополагающих положений теории чисел здесь строго доказан, и эти доказательства в течение уже более двух тысячелетий являются образцами математических доказательств.

✓*Евклид*

Дошедшие до нас сведения о Евклиде, чьи созданные около 300 года до н. э. в Александрии «Начала» на два тысячелетия стали кладезем математической премудрости, чрезвычайно скудны. Известен ставший хрестоматийным ответ Евклида египетскому царю Птолемею I, что в геометрии нет царских дорог. Античные источники передают также ещё один рассказ о Евклиде. Начавший заниматься у Евклида ученик спросил его, что он сможет заработать, выучив всё это. Тогда Евклид позвал своего раба и сказал: «Дай ему три монеты, ибо он должен добывать тем, что учит».

О высоком уровне развития теории чисел и значении, придаваемом искусству в этой области в эпоху эллинизма, свидетельствует знаменитая задача Архимеда о быках. Величайший математик и механик древности Архимед (287–212 до н. э.), живший в Сиракузах на Сицилии, обессмертил своё имя открытием закона рычага, закона Архимеда в гидростатике, а также вычислением площадей и объёмов ряда фигур методами, предвосхитившими современное интегральное исчисление. Найденная Архимедом формула для объёма шара $V = \frac{4}{3}\pi R^3$ была согласно его завещанию изображена в виде вписанного в цилиндр шара на его надгробии, а высеченная под рисунком эпитафия сообщала, что объёмы этих тел относятся как 3 : 2. Сто тридцать семь лет спустя после смерти Архимеда этот памятник среди терниев и чертополоха ещё удалось отыскать Цицерону. Время не сохранило до нас надгробия Архимеда, погибшего при штурме Сиракуз римлянами в 212 году до н. э., но сама формула является вечным памятником её первооткрывателю. Знаменитое же восклицание Архимеда «Эврика!» (я нашёл!), сделанное им согласно легенде при установлении закона гидростатики, стало символом научного открытия.

✓*Архимед*

Задача о быках ➡

Математический гений Архимеда был, воистину, универсален. Об этом свидетельствует, в частности, его знаменитая «Задача о быках», предложенная им своим александрийским коллегам в письме, адресованном главе Александрийской библиотеки Эратосфену. В задаче, написанной эпическим ионийским стихом, предлагается найти число быков и коров в каждом из четырёх стад — белом, чёрном, пёстром и буром, принадлежащих богу Солнца (Аполлону-Гелиосу) и пасущихся на сочных лугах Сицилии:[1]

[1] Стихи цитируются по книге: *Лурье С. Я.* Архимед. М.—Л.: Издательство Академии Наук СССР, 1945.

Сколько у Солнца коров и быков, сосчитай, чужестранец
Ум навостривши, коль впрямь свойственна мудрость тебе.
Сколько скота выгонялось на долы Сицилии влажной?
Разного цвета стада бог лучезарный имел,
Счётом четыре: из них одно — белоснежное было,
Чёрным отливом других лоснилась жирная шерсть,
Бурое — третье стадо, четвёртое — пёстрое.

Далее также в стихотворной форме формулируются условия задачи. Если обозначить число быков в белом, чёрном, пёстром и буром стадах через x, y, z и t, а число коров — соответственно через x', y', z' и t', то в современной символике эти условия запишутся в виде системы из 7 уравнений с 8 неизвестными, решить которую нужно в натуральных числах:

$$\begin{cases} x = (\frac{1}{2}+\frac{1}{3})y + t, \\ y = (\frac{1}{4}+\frac{1}{5})z + t, \\ z = (\frac{1}{6}+\frac{1}{7})x + t, \\ x' = (\frac{1}{3}+\frac{1}{4})(y + y'), \\ y' = (\frac{1}{4}+\frac{1}{5})(z + z'), \\ z' = (\frac{1}{5}+\frac{1}{6})(t + t'), \\ t' = (\frac{1}{6}+\frac{1}{7})(x + x') \ . \end{cases}$$

Исследование этой системы показывает, что она имеет бесконечное множество решений, наименьшее из которых есть $x_0 = 10366482$, $y_0 = 7460514$, $z_0 = 7358060$, $t_0 = 4149387$, $x'_0 = 7206360$, $y'_0 = 4893246$, $z'_0 = 3515820$, $t'_0 = 5439213$, а все остальные ему кратны: $(x, y, z, t, x', y', z', t') = (kx_0, ky_0, kz_0, kt_0, kx'_0, ky'_0, kz'_0, kt'_0)$, где $k \in \mathbb{N}$.

В таком виде задача была вполне по силам александрийским математикам того времени. Решивший эту задачу, по мнению Архимеда, может считать себя разбирающимся в числах, но не более того:

Если сочтёшь скота всего там сколько набралось,
Сколько паслось на лугах мясообильных быков,
Сколько удойных коров и сколько каждого цвета
Не назовёт уж никто в числах невеждой тебя.
Всё же и к мудрым ещё тебя не причислят за это,
Коль не учтёшь ты ещё разных повадок быков...

«Повадки» быков оказываются весьма коварными. В математической форме они выражаются в том, что $x + y$ должно было быть квадратным

числом, а $z+t$ — треугольным. Дополнительные условия сильно усложняют задачу, и Архимед продолжает далее:

Если сумеешь всё это найти и взором духовным
Стада размеры объять сам и другим передать,
Гордо шествуй вперёд, кичася великой победой:
Знай, что, других превзойдя, первый по мудрости ты.

Смысл сквозящей в последних строках иронии становится понятен, если принять во внимание, что приблизиться к решению этой задачи удалось лишь в конце XIX века, когда выяснилось, что общее поголовье скота в четырёх стадах выражается десятичным числом с 206 545 знаками, требующим для своей записи более полусотни страниц убористого книжного текста. Точное же решение удалось получить с помощью компьютера лишь к концу XX века.

✓*Диофант*

Позже, задачи, связанные с решением неопределённых уравнений в целых числах, систематически рассматривались Диофантом (III век н. э.) в его «Арифметике». Подобные уравнения теперь и называют диофантовыми. Диофант был последним великим математиком античности. Дальнейшее продвижение в теории чисел после почти полуторатысячелетнего застоя связано уже с именем Пьера Ферма, исходной точкой для которого были исследования Диофанта. В заметках Ферма на полях «Арифметики» Диофанта 1621 года издания и появилась впервые его знаменитая Великая теорема: для любого натурального $n>2$ уравнение $x^n+y^n=z^n$ не имеет целых положительных решений. Хотя Ферма и записал на полях, что располагает доказательством этой теоремы, доказательство этого, возможно, самого знаменитого в истории математики результата, удалось найти лишь в самом конце двадцатого века. Причём при его доказательстве английским математиком Эндрю Уайлсом были использованы последние достижения современной алгебраической геометрии.

У Диофанта же впервые появляются и элементы алгебраической символики, усовершенствованные позже Франсуа Виетом (1540–1603) и Рене Декартом. О самом Диофанте, кроме того, что он жил в Александрии, неизвестно практически ничего. Даже годы его жизни удалось установить по косвенным свидетельствам с точностью лишь до века. Одним из немногих источников информации о его жизни является дошедшая до нашего времени античная эпитафия-загадка:

Прах Диофанта гробница покоит: дивись ей — и камень
Мудрым искусством его скажет усопшего век.
Волей богов шестую часть жизни он прожил ребёнком,
И половину шестой встретил с пушком на щеках.
Только минула седьмая, с подругою он обручился.
С нею пять лет проведя, сына дождался мудрец.
Только полжизни отцовской возлюбленный сын его прожил.
Отнят он был у отца ранней могилой своей.
Дважды два года родитель оплакивал тяжкое горе.
Тут и увидел предел жизни печальной своей.

Читателю предоставляется возможность самому решить, сколько лет прожил Диофант.

В течение тысячелетий теория чисел являла собой область чистой математики, не предполагавшей какого-либо практического использования. Многие выдающиеся математики прошлого, включая великого немецкого математика Гаусса (1777–1855), считали её «царицей математики», по самой своей природе свободной от любых приложений. Положение, однако, кардинально изменилось во второй половине XX века, когда для неё нашлись важные приложения, сначала в программах датчиков случайных чисел, а затем в кодировании и криптографии. Знакомство с основами теории чисел необходимо при изучении дискретной математики.

◆ Деление с остатком

Вопросы делимости являются центральными в теории чисел. Натуральное число d делит натуральное число n, если существует такое натуральное число k, что $n = k\,d$. Это записывается как $d \mid n$ и говорят, что d является *делителем* n, а n — *кратным* d. Число же k называется *частным* от деления n на d.

Такое деление не всегда выполнимо во множестве натуральных чисел. Однако всегда выполнимо деление на меньшее число с остатком: всякое натуральное n представимо единственным образом через меньшее его натуральное число m в форме

$$n = k\,m + r, \text{ где } k, r \in N, \text{ причём } 0 \le r < m.$$

Число r называется *остатком* от деления n на m. Равенство остатка нулю означает, что $m \mid n$. Таким образом, при делении пяти на три получаем 1 в частном и 2 в остатке: $5 = 1 \cdot 3 + 2$.

Чтобы узнать, сколько среди первых N натуральных чисел таких, которые делятся на m, достаточно разделить N на m с остатком. Частное от этого деления и даёт ответ на поставленный вопрос. Так, среди первой сотни натуральных чисел четырнадцать делятся на 7, так как $100 = 14 \cdot 7 + 2$.

◆ Простые числа

Каждое натуральное число n, большее единицы, имеет, по крайней мере, два делителя: 1 и n. Особую роль в теории играют те натуральные числа, которые не имеют других делителей кроме этих двух. Они называются *простыми*. Вот начальный отрезок бесконечной, как доказано у Евклида, последовательности простых чисел: 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, ... Натуральные числа, большие единицы и не являющиеся простыми, называются *составными*. Единица занимает особое положение в ряду натуральных чисел и не относится ни к простым, ни к составным числам.

Как часто среди натуральных чисел встречаются простые? Рассмотрение приведённого начального отрезка последовательности простых чисел

позволяет заметить в ней *близнецов* — пар простых чисел, разность между которыми равна 2, например, 11 и 13 или 29 и 31. С помощью компьютера были найдены весьма большие близнецы, но до сих пор неизвестно конечно или бесконечно множество близнецов, и это — одна из самых интригующих проблем в теории чисел. С другой стороны в натуральном ряду существуют сколь угодно длинные отрезки, лишённые простых чисел. Например, $n-1$ последовательных чисел $n!+2, n!+3, \ldots, n!+n$ делятся, соответственно, на $2,3, \ldots, n$ и поэтому являются составными.

Несмотря на ряд трудных и до сих пор нерешённых задач, связанных с распределением простых чисел, частота их встречаемости, тем не менее, установлена точно. Число простых чисел, не превосходящих n, принято обозначать через $\pi(n)$. К концу XIX века усилиями многих математиков, начиная с Гаусса, было показано, что

$$\lim_{n\to\infty} \frac{\pi(n)/n}{1/\ln n} = 1.$$

Разложение на множители ➡

Допуская некоторую вольность, можно сказать, что взятое наугад большое натуральное число n с вероятностью $1/\ln n$ будет простым. Каждое натуральное число, большее единицы, единственным образом раскладывается в произведение простых чисел, например, $360 = 2^3 \cdot 3^2 \cdot 5$. Простые числа являются, таким образом, теми «атомами», из которых с помощью операции умножения строится всё множество натуральных чисел. Этот хорошо знакомый каждому с начальных классов средней школы факт воспринимается настолько естественно, что кажется не требующим доказательства. Однако при более критическом рассмотрении становится отнюдь не очевидным, почему произведение некоторого конечного множества простых чисел с заданными кратностями их повторений не может совпасть с произведением другого подобного множества. Единственность разложения на простые множители является важнейшим свойством натуральных чисел. Доказательство этого факта, который иногда называют основной теоремой арифметики, приведено в конце этой главы.

Общие делители и кратные ➡

Натуральное d, делящее одновременно натуральные m и n, называется их *общим делителем*. *Наибольший общий делитель* чисел m и n обозначается (m, n). Из единственности разложения на простые множители следует, что наибольший общий делитель (m, n) можно найти, разложив числа m и n на простые множители и составив затем произведение из степеней простых чисел, встречающихся в этих разложениях, выбрав степень каждого из них по его меньшей степени в двух разложениях. Пусть, например, $m = 60$, $n = 72$. Имеем $60 = 2^2 \cdot 3^1 \cdot 5^1$, $72 = 2^3 \cdot 3^2 \cdot 5^0$.

Поэтому $(60,72)=2^2\cdot 3^1\cdot 5^0=12$. Каждый общий делитель чисел m и n делит (m, n).

Натуральное число, которое кратно как m, так и n, называется их *общим кратным*. *Наименьшее общее кратное* чисел m и n обозначается $[m, n]$. Наименьшее общее кратное также может быть найдено через разложение на простые множители, но их степени теперь должны выбираться по наибольшей степени данного множителя в каждом из двух разложений. Например, если $m=60$, $n=72$, то $[60,72]=2^3\cdot 3^2\cdot 5^1=360$. Каждое общее кратное чисел m и n кратно $[m, n]$.

Два натуральных числа m и n называются *взаимно простыми*, если $(m, n)=1$, т. е. единица является их единственным общим делителем. В качестве примера можно привести, например, числа 9 и 14 или 16 и 21. Простое число будет, разумеется, взаимно простым с любым отличным от него натуральным числом.

Для данного натурального n число натуральных чисел, не превышающих n и взаимно простых с n, называется *функцией Эйлера* и обозначается $\varphi(n)$. Имеем $\varphi(1)=1$, $\varphi(2)=1$, $\varphi(3)=2$, $\varphi(4)=2$, $\varphi(5)=4$, $\varphi(6)=2$, $\varphi(7)=6$ и т. д. Ясно, что если p — простое, то $\varphi(p)=p-1$.

← Функция Эйлера

Для всех натуральных n функция Эйлера удовлетворяет красивому тождеству, впервые подмеченному Гауссом,

$$\sum_{d|n}\varphi(d)=n,$$

где суммирование в левой части распространяется на все делители n. Например, если $n=6$, то

$$\varphi(1)+\varphi(2)+\varphi(3)+\varphi(4)=1+1+2+2=6.$$

Для доказательства тождества достаточно рассмотреть n дробей

$$\frac{1}{n},\frac{2}{n},\frac{3}{n},\ldots,\frac{n}{n}$$

и произвести их сокращение так, чтобы числитель и знаменатель каждой дроби стали взаимно простыми. Тогда для каждого $d\mid n$ возникнут, как нетрудно понять, все $\varphi(d)$ дробей вида e/d, где $e\le d$ и $(e,d)=1$. Например, для $n=6$ получаем

$$\frac{1}{6},\frac{2}{6},\frac{3}{6},\frac{4}{6},\frac{5}{6},\frac{6}{6}\Rightarrow\frac{1}{6},\frac{1}{3},\frac{1}{2},\frac{2}{3},\frac{5}{6},\frac{1}{1}.$$

Из этого замечания и вытекает справедливость тождества.

Завершая этот краткий экскурс в теорию чисел, вернёмся к совершенным числам и докажем предложение 36 книги IX «Начал» Евклида,

Формула совершенного числа ➡

которое утверждает, что если сумма $1+2+2^2+\ldots+2^n=2^{n+1}-1=p$ — есть число простое, то $2^n p$ — совершенное число. В самом деле, собственными делителями числа $2^n p$ являются $1,2,2^2,\ldots,2^n$ и $p, 2p, 2^2 p, \ldots, 2^{n-1} p$, и их сумма есть $p+(2^n-1)p=2^n p$.

Первыми подходящими значениями n являются $1,2,4,6,12$, что даёт совершенные числа 6, 28, 496, 8128, 33 550 336, первые четыре из которых были известны пифагорейцам. А есть ли совершенные числа отличного от $2^n p$ вида или задолго до нашей эры пифагорейцы сумели описать всё их множество? В XVIII веке Эйлер доказал, что чётных совершенных чисел другого вида не существует. А есть ли нечётные совершенные числа? Этот вопрос уже третье тысячелетие остаётся открытым. Вообще, в вопросе о совершенных числах математической науке не удалось за две с половиной тысячи лет продвинуться существенно дальше древних греков. К концу XX века благодаря компьютеру количество известных совершенных чисел достигло 27, и наибольшее из них равно $2^{44\,496}(2^{44\,497}-1)$. По-прежнему, однако, остаётся неизвестным, конечно или бесконечно их число.

Векторы. *Векторы* появляются практически во всех разделах современной математики. Множество векторов с операцией сложения векторов и умножения их на действительные числа называется *векторным пространством*. Простейшей моделью векторного пространства является множество векторов на плоскости (рис. 15).

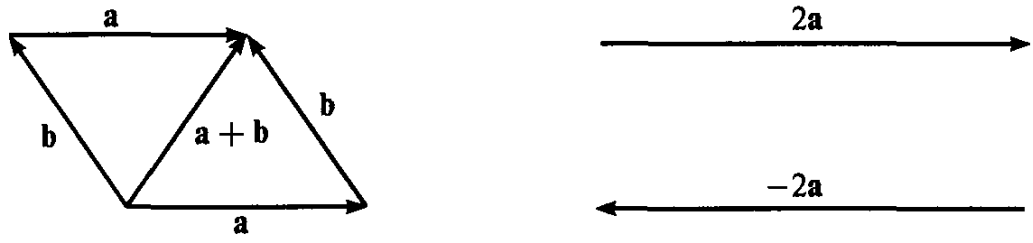


Рис. 15

Действия над векторами ➡

В общем случае векторы могут быть элементами любой природы с определёнными на них операциями сложения и умножения на число со следующими свойствами:

1) $\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}$ (коммутативность сложения);
2) $(\mathbf{a}+\mathbf{b})+\mathbf{c}=\mathbf{a}+(\mathbf{b}+\mathbf{c})$ (ассоциативность сложения);
3) имеется нуль-вектор $\mathbf{0}$ такой, что $\mathbf{a}+\mathbf{0}=\mathbf{a}$ для любого $\mathbf{a}$;

4) для любого вектора $\mathbf{a}$ существует противоположный ему вектор $-\mathbf{a}$ такой, что $\mathbf{a}+(-\mathbf{a})=\mathbf{0}$ (при этом прибавление противоположного вектора записывается как вычитание векторов $\mathbf{a}+(-\mathbf{b})=\mathbf{a}-\mathbf{b}$);
5) $1\cdot\mathbf{a}=\mathbf{a}$ и $0\cdot\mathbf{a}=\mathbf{0}$;
6) $\alpha(\beta\mathbf{a})=(\alpha\beta)\mathbf{a}$;
7) $(\alpha+\beta)\mathbf{a}=\alpha\mathbf{a}+\beta\mathbf{a}$;
8) $\alpha(\mathbf{a}+\mathbf{b})=\alpha\mathbf{a}+\alpha\mathbf{b}$.

Выражение $\alpha_1\mathbf{a}_1+\alpha_2\mathbf{a}_2+\ldots+\alpha_k\mathbf{a}_k$ называется *линейной комбинацией* векторов $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ с коэффициентами $\alpha_1, \alpha_2, \ldots, \alpha_k$. Множество векторов, порождённых всевозможными линейными комбинациями векторов $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$, называется их *линейной оболочкой*. Линейная комбинация называется *нетривиальной*, если хотя бы один из коэффициентов $\alpha_1, \alpha_2, \ldots, \alpha_k$ отличен от нуля. Векторы $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ называются *линейно зависимыми*, если существует обращающаяся в нуль-вектор их нетривиальная линейная комбинация, и *линейно независимыми* — в противном случае.

Линейная зависимость ⬅ и независимость

Линейная зависимость эквивалентна тому, что хотя бы один из векторов является линейной комбинацией остальных. В самом деле, если $\alpha_1\mathbf{a}_1+\alpha_2\mathbf{a}_2+\ldots+\alpha_k\mathbf{a}_k=\mathbf{0}$ и $\alpha_k\neq 0$, то $\mathbf{a}_k=-\frac{\alpha_1}{\alpha_k}\mathbf{a}_1-\frac{\alpha_2}{\alpha_k}\mathbf{a}_2-\ldots-\frac{\alpha_{k-1}}{\alpha_k}\mathbf{a}_k$.

Два вектора на плоскости линейно зависимы в том и только в том случае, если они коллинеарны (направлены вдоль одной прямой). Любые три вектора на плоскости всегда линейно зависимы, так как любой вектор на плоскости может быть представлен в виде линейной комбинации двух заданных неколлинеарных векторов (рис. 16).

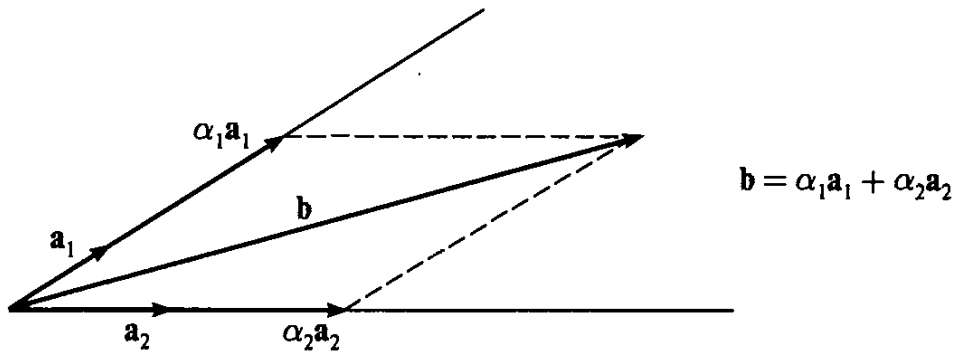


Рис. 16

В пространстве три вектора зависимы в том и только в том случае, если они компланарны (параллельны заданной плоскости). Любые четыре вектора в пространстве линейно зависимы, так как любой вектор может

быть выражен в виде линейной комбинации трёх заданных некомпланарных векторов.

Размерность и базис ➡

Число n называется размерностью $\dim V$ векторного пространства V, если в V существуют n линейно независимых векторов, а любые $n+1$ векторов линейно зависимы. Размерность плоскости равна, таким образом, двум, а размерность пространства — трём. В математике рассматриваются векторные пространства сколь угодно большой и даже бесконечной размерности, но в дискретной математике пространства бесконечной размерности не возникают. Любые n линейно независимых векторов n-мерного векторного пространства образуют *базис* этого пространства. Если $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ — базис, то любой вектор $\mathbf{a}$ может быть представлен в виде линейной комбинации базисных векторов $\mathbf{a} = \alpha_1\mathbf{e}_1 + \alpha_2\mathbf{e}_2 + \ldots + \alpha_n\mathbf{e}_n$, что сразу следует из линейной зависимости системы векторов $\mathbf{a}, \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$. Это представление единственно, так как если $\mathbf{a} = \alpha_1'\mathbf{e}_1 + \alpha_2'\mathbf{e}_2 + \ldots + \alpha_n'\mathbf{e}_n$, то $(\alpha_1 - \alpha_1')\mathbf{e}_1 + (\alpha_2 - \alpha_2')\mathbf{e}_2 + \ldots + (\alpha_n - \alpha_n')\mathbf{e}_n = \mathbf{0}$, и по свойству базиса $\alpha_1 = \alpha_1'$, $\alpha_2 = \alpha_2'$, ..., $\alpha_n = \alpha_n'$. Коэффициенты $\alpha_1, \alpha_2, \ldots, \alpha_n$ называются координатами вектора $\mathbf{a}$ в базисе $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$.

В качестве примера векторного пространства размерности n можно рассмотреть множество всех многочленов $a_0 + a_1x + a_2x^2 + \ldots + a_{n-1}x^{n-1}$, степени не превышающей $n-1$, с обычными алгебраическими операциями сложения многочленов и умножения их на действительные числа. Естественным базисом в этом пространстве являются многочлены $1, x, x^2, \ldots, x^{n-1}$.

Если в n-мерном векторном пространстве задан некоторый базис, то тем самым установлено взаимно однозначное соответствие между векторами и наборами их координат в этом базисе $\mathbf{a} \leftrightarrow (a_1, a_2, \ldots, a_n)$. При этом $\lambda\mathbf{a} \leftrightarrow (\lambda a_1, \lambda a_2, \ldots, \lambda a_n)$ и $\mathbf{a} + \mathbf{b} \leftrightarrow (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n)$, т. е. векторные операции выполняются покоординатно.

Само множество наборов действительных чисел $(a_1, a_2, \ldots, a_n)$ также можно считать векторным пространством, векторные операции в котором выполняются согласно правилам

$$(a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n) = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n),$$
$$\lambda(a_1, a_2, \ldots, a_n) = (\lambda a_1, \lambda a_2, \ldots, \lambda a_n).$$

Такое векторное пространство называется n-мерным *арифметическим пространством*. Векторы $(100\ldots0)$, $(010\ldots0)$, ..., $(000\ldots1)$ образуют в нём естественный базис. Компоненты a_i, $i = 1, 2, \ldots, n$, вектора $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ являются его координатами в этом базисе.

Помимо сложения векторов и умножения их на действительные числа векторное пространство может быть наделено и некоторыми другими, до-

полнительными операциями, важнейшей из которых является *скалярное произведение*, ставящее в соответствие каждой паре векторов действительное число. Векторное пространство, наделённое скалярным произведением, называется *евклидовым векторным пространством*. На плоскости скалярное произведение $(\mathbf{a},\mathbf{b})$ векторов $\mathbf{a}$ и $\mathbf{b}$ есть произведение их длин, умноженное на косинус угла между ними: $(\mathbf{a},\mathbf{b}) = |\mathbf{a}| \cdot |\mathbf{b}| \cdot \cos\alpha$. Отсюда $(\mathbf{a},\mathbf{a}) = |\mathbf{a}|^2$, а если $\mathbf{a} \perp \mathbf{b}$, то $(\mathbf{a},\mathbf{b}) = 0$. В общем случае для скалярного произведения в евклидовом пространстве требуется выполнение следующих свойств:

Скалярное произведение

1) $(\mathbf{a},\mathbf{b}) = (\mathbf{b},\mathbf{a})$ (коммутативность);
2) $(\mathbf{a}+\mathbf{b},\mathbf{c}) = (\mathbf{a},\mathbf{c}) + (\mathbf{b},\mathbf{c})$ (дистрибутивность);
3) $(\lambda\mathbf{a},\mathbf{b}) = \lambda(\mathbf{a},\mathbf{b})$;
4) $(\mathbf{a},\mathbf{a}) > 0$ для всех $\mathbf{a} \neq \mathbf{0}$.

Евклидовы пространства обладают многими привычными свойствами векторов на плоскости и в пространстве. Если $(\mathbf{a},\mathbf{b}) = 0$, то векторы $\mathbf{a}$ и $\mathbf{b}$ называются *ортогональными*, что для векторов на плоскости и в пространстве означает перпендикулярность. Базис $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ называется ортонормированным, если $(\mathbf{e}_i, \mathbf{e}_i) = 1$, $i = 1, 2, \ldots, n$ и $(\mathbf{e}_i, \mathbf{e}_j) = 0$ при $i \neq j$. Если $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ — ортонормированный базис, $\mathbf{a} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \ldots + a_n\mathbf{e}_n$, $\mathbf{b} = b_1\mathbf{e}_1 + b_2\mathbf{e}_2 + \ldots + b_n\mathbf{e}_n$, то, как следует из свойств скалярного произведения, $(\mathbf{a},\mathbf{b}) = a_1b_1 + a_2b_2 + \ldots + a_nb_n$. Это выражение принимается за определение скалярного произведения в евклидовом n-мерном арифметическом пространстве.

Ортогональность

Подобно тому как на плоскости существуют пары ортогональных векторов, а в пространстве — тройки попарно ортогональных векторов, в евклидовом n-мерном пространстве всегда существует ортогональный базис. В самом деле, если $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n$ — произвольный базис, то ортогональный базис $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ может быть получен в результате следующего процесса ортогонализации. Полаем $\mathbf{e}_1 = \mathbf{f}_1$. Вектор $\mathbf{e}_2$ ищем в виде $\mathbf{e}_2 = \mathbf{f}_2 + \alpha\mathbf{e}_1$. Из условия $(\mathbf{e}_1,\mathbf{e}_2) = 0$ получаем $\alpha = -(\mathbf{f}_2,\mathbf{e}_1)/(\mathbf{e}_1,\mathbf{e}_1)$ (рис. 17).

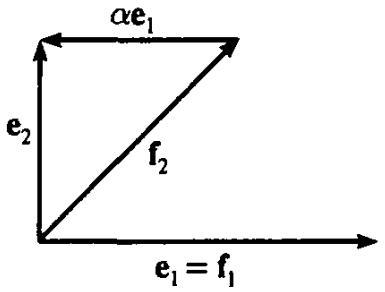


Рис. 17

Вектор $\mathbf{e}_3$ ищем в виде $\mathbf{e}_3 = \mathbf{f}_3 + \beta_1\mathbf{e}_1 + \beta_2\mathbf{e}_2$. Из условия $(\mathbf{e}_1,\mathbf{e}_3) = 0$ получаем $\beta_1 = -(\mathbf{f}_3,\mathbf{e}_1)/(\mathbf{e}_1,\mathbf{e}_1)$, а из условия $(\mathbf{e}_2,\mathbf{e}_3) = 0$ — $\beta_2 = -(\mathbf{f}_3,\mathbf{e}_2)/(\mathbf{e}_2,\mathbf{e}_2)$. Четвёртый базисный вектор ищем в виде $\mathbf{e}_4 = \mathbf{f}_4 + \gamma_1\mathbf{e}_1 + \gamma_2\mathbf{e}_2 + \gamma_3\mathbf{e}_3$ и т. д.

В результате получаем ортогональный базис $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$, который может быть преобразован в ортонормированный нормировкой базисных векторов $\mathbf{e}'_i = \mathbf{e}_i/(\mathbf{e}_i, \mathbf{e}_i)$, $i = 1, 2, \ldots, n$.

Подпространства ➡

Подпространством векторного пространства называется всякое подмножество его векторов, само образующее векторное пространство, т. е. замкнутое относительно операций сложения векторов и умножения их на числа. Множество коллинеарных векторов на плоскости — подпространство размерности 1, множество компланарных векторов в пространстве — подпространство размерности 2. При $k < n$ пространство многочленов, степень которых не превышает k, является подпространством пространства многочленов, степень которых не превышает n.

Если S и T два подпространства векторного пространства V, то их пересечение $S \cap T$ и сумма $S + T$ также являются подпространствами пространства V. При этом под суммой $S + T$ понимается множество всех векторов вида $\mathbf{s} + \mathbf{t}$, где $\mathbf{s} \in S$, $\mathbf{t} \in T$. Размерности этих подпространств связаны соотношением $\dim(S + T) = \dim S + \dim T - \dim(S \cap T)$. Чтобы убедиться в этом, достаточно выбрать базис в $S \cap T$ и дополнить его, с одной стороны, до базиса в S, с другой — до базиса в T. Полученные $\dim S + \dim T - \dim(S \cap T)$ векторов будут линейно независимы и любой вектор из $S + T$ будет выражаться их линейной комбинацией, поэтому они образуют базис в $S + T$.

Ортогональное дополнение ➡

В евклидовом пространстве V множество векторов, ортогональных всем векторам некоторого подпространства U, образует подпространство $U^\perp$, которое называется *ортогональным дополнением* подпространства U. Отметим, прежде всего, что, как следует из свойства 4) скалярного произведения, единственным общим элементом подпространств U и $U^\perp$ является нуль-вектор. Далее, сумма размерностей U и $U^\perp$ равна размерности всего пространства V: $\dim U + \dim U^\perp = \dim V$. Чтобы убедиться в этом, достаточно выбрать ортогональный базис в U и дополнить его до ортогонального базиса в V. Дополненные векторы образуют базис в $U^\perp$. Отсюда также следует, что каждый вектор $\mathbf{v} \in V$ может быть разложен в сумму $\mathbf{v} = \mathbf{u} + \mathbf{u}'$, где $\mathbf{u} \in U$, $\mathbf{u}' \in U^\perp$. Это разложение единственно, так как если $\mathbf{v} = \mathbf{u}_1 + \mathbf{u}'_1$, то $\mathbf{u} - \mathbf{u}_1 = \mathbf{u}'_1 - \mathbf{u}'$ и $(\mathbf{u} - \mathbf{u}_1, \mathbf{u}'_1 - \mathbf{u}') = 0$, откуда $\mathbf{u} = \mathbf{u}_1$ и $\mathbf{u}'_1 = \mathbf{u}'$. Все эти свойства кратко выражают, говоря, что V разлагается в *прямую сумму* подпространств U и $U^\perp$.

В качестве примера рассмотрим в пространстве множество векторов, компланарных заданной плоскости, которое образует подпространство

размерности 2. Ортогональное дополнение к нему состоит из векторов, перпендикулярных этой плоскости, и является одномерным подпространством. При этом каждый вектор пространства является суммой своих проекций на плоскость и на ось, перпендикулярную плоскости, и это представление единственно.

Другой пример, множество решений линейного однородного уравнения $a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0$, рассматриваемое как множество векторов $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ в n-мерном арифметическом пространстве, является $(n-1)$-мерным подпространством, ортогональным одномерному подпространству, порождённому вектором $\mathbf{a} = (a_1, a_2, \ldots, a_n)$.

Системы
⬅ уравнений

Легко проверить также следующее свойство ортогональных дополнений: $(S+T)^{\perp} = S^{\perp} \cap T^{\perp}$ (задача 30), которое распространяется на любое число подпространств: $(U_1 + U_2 + \ldots + U_k)^{\perp} = U_1^{\perp} \cap U_2^{\perp} \cap \ldots \cap U_k^{\perp}$. В качестве примера рассмотрим множество решений системы линейных однородных уравнений

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots a_{1n}x_n = 0, \\ a_{21}x_1 + a_{22}x_2 + \ldots a_{2n}x_n = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots a_{mn}x_n = 0. \end{cases}$$

Положив $\mathbf{a}_1 = (a_{11}, a_{12}, \ldots, a_{1n})$, $\mathbf{a}_2 = (a_{21}, a_{22}, \ldots, a_{2n})$, ..., $\mathbf{a}_m = (a_{m1}, a_{m2}, \ldots, a_{mn})$, систему можно переписать в виде $(\mathbf{a}_i, \mathbf{x}_i) = 0$, $i = 1, 2, \ldots, m$. Обозначив через A_1, A_2, ..., A_m одномерные подпространства, порождённые соответственно векторами $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$, для множества X решений системы получаем

$$X = A_1^{\perp} \cap A_2^{\perp} \cap \ldots \cap A_m^{\perp} = (A_1 + A_2 + \ldots + A_m)^{\perp}.$$

Множество $A_1 + A_2 + \ldots + A_m$ есть линейная оболочка векторов $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$. Размерность линейной оболочки системы векторов $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$, равная максимальному числу линейно независимых среди них, называется *рангом* r этой системы. Если векторы $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ линейно независимы, то их ранг $r = m$, в противном случае $r < m$. Как следует из вышеизложенного, размерность пространства решений есть $\dim X = n - r$.

Отношения. Множество, лишённое каких-либо отношений между своими элементами, является простым собранием своих членов и полностью характеризуется своей мощностью. Гораздо более содержательными со всех точек зрения объектами являются множества, наделённые определёнными отношениями. Если взять некоторый коллектив людей и рассмотреть упо-

рядоченную пару (a,b), взятую из этого множества, то члены пары могут находиться в некотором отношении R, что обозначается, как $(a,b) \in R$. В качестве такого отношения R может выступать, например, отношение знакомства, тогда $(a,b) \in R$ означает, что a и b знакомы между собой. В этом случае, если $(a,b) \in R$, то и $(b,a) \in R$, что выражает симметричность отношения знакомства. Если же в качестве отношения R рассмотреть долговое отношение, когда $(a,b) \in R$ означает, что a должен b некоторую сумму денег, то из $(a,b) \in R$ следует $(b,a) \notin R$, что выражает асимметричность долгового отношения.

Подобные бинарные отношения естественным образом возникают и на числовых множествах. В качестве примеров можно привести хорошо известные из арифметики отношения «=» (равно), «<» (меньше), «≤» (меньше или равно), «>» (больше), «≥» (больше или равно). Здесь отношение «=» является примером симметричного отношения, а отношения «<» и «>» — примерами асимметричных отношений. Рассмотрев подобные поясняющие примеры, определим математическое понятие бинарного отношения в самом общем виде.

Пусть A — произвольное множество, $A^2 = A \times A$ — его декартов квадрат (множество всех упорядоченных пар его элементов). Если задано некоторое подмножество $R \subseteq A^2$, т. е. некоторое множество упорядоченных пар (a_i, a_j), где $a_i, a_j \in A$, то говорят, что на множестве A задано *бинарное отношение* R. Пишут $(a_i, a_j) \in R$ или $a_i R a_j$.

Для каждого отношения R обратное ему отношение R^{-1} определяется следующим образом: $(a_i, a_j) \in R^{-1}$ тогда и только тогда, когда $(a_j, a_i) \in R$. Отношения «<» и «>», а также «≤» и «≥» являются примерами пар взаимно обратных отношений.

Свойства бинарных отношений ➡

Выделяют следующие основные свойства бинарных отношений.

1. *Рефлексивность* — для любого $a \in A$ имеет место $(a,a) \in R$, т. е. каждый элемент множества A находится в отношении R сам с собой.
2. *Иррефлексивность* — для любого $a \in A$ имеет место $(a,a) \notin R$, т. е. никакой элемент множества A не находится в отношении R сам с собой.
3. *Симметричность* — из $(a,b) \in R$ следует $(b,a) \in R$, т. е. отношение R не зависит от порядка, в котором берутся элементы a и b.
4. *Асимметричность* — из $(a,b) \in R$ следует $(b,a) \notin R$.
5. *Антисимметричность* — из $(a,b) \in R$ и $(b,a) \in R$ следует $a = b$.
6. *Транзитивность* — из $(a,b) \in R$ и $(b,c) \in R$ следует $(a,c) \in R$.

Таким образом, отношение «=» рефлексивно, симметрично и транзитивно. Отношения «<» и «>» иррефлексивны, асимметричны и транзитивны. Отношения «≤» и «≥» рефлексивны, антисимметричны и транзитивны.

Отношение порядка

Два вида отношений, характеризующиеся определёнными наборами перечисленных свойств, являются особенно важными — это *отношение порядка* и *отношение эквивалентности*.

Бинарное отношение R на множестве A, являющееся одновременно рефлексивным, антисимметричным и транзитивным, называется отношением порядка.

Таким образом, отношения «≤» и «≥» являются отношениями порядка.

Пусть R — отношение порядка. Если $(a,b) \in R$, то говорят, что элемент a предшествует элементу b или элемент b следует за элементом a, а отношение aRb часто пишут как $a \le b$, сохраняя привычный арифметический символ. Запись $b \ge a$ обозначает обратное отношение, т. е. эквивалентна записи $a \le b$. Запись $a < b$ означает, что $a \le b$ и $a \ne b$.

Если для любых двух элементов $a,b \in A$ имеет место хотя бы одно из отношений $a \le b$ или $b \le a$, то порядок называется *полным* или *линейным* порядком, а множество A — *линейно упорядоченным* или *цепью*. Обычное отношение «≤» на множестве действительных чисел и его подмножествах является примером линейного порядка.

Система подмножеств некоторого множества, упорядоченная по включению: $A \le B \Leftrightarrow A \subseteq B$, даёт пример порядка, не являющегося линейным. В таком случае говорят о *частично упорядоченном множестве*.

Другим интересным примером частично упорядоченного множества является множество натуральных чисел, упорядоченное отношением делимости. В этом порядке $m \le n$, если $m \mid n$. Так $3 \le 6$, но неверно, что $4 \le 6$.

В качестве ещё одного примера частичного порядка можно привести множество разбиений некоторого множества. Здесь разбиение β следует за разбиением α $(\alpha \le \beta)$, если α получается измельчением β, т. е. разбиением некоторых подмножеств разбиения β. Например, если рассматриваются разбиения множества $\{1,2,3,4,5\}$ и $\alpha = \{1,2,3\} \cup \{4\} \cup \{5\}$, а $\beta = \{1,2,3\} \cup \{4,5\}$, то $\alpha \le \beta$.

Элемент a упорядоченного множества называется *минимальным*, если из $b \le a$ следует $b = a$. Упорядоченное множество может не иметь ни одного минимального элемента, как, например, множество целых чисел $Z = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ с обычным отношением порядка между числами. Но подобная ситуация возможна лишь для бесконечных множеств. Конечные множества обязательно имеют минимальные элементы, которых может быть один или больше. Так, в системе из пяти множеств

$\{a\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}$, упорядоченных по включению, имеется два минимальных элемента: $\{a\}$ и $\{b,c\}$. В системе из семи натуральных чисел $\{2,3,4,6,8,12,16\}$, упорядоченных отношением делимости, также имеется два минимальных элемента: 2 и 3.

Элемент a называется *наименьшим*, если для любого b имеет место $a \le b$. Если в упорядоченном множестве имеется наименьший элемент, то он единственен. В самом деле, если a и b два наименьших элемента, то $a \le b$ и $b \le a$ и по свойству антисимметричности имеем $a = b$. Ясно также, что если в конечном частично упорядоченном множестве имеется единственный минимальный элемент, то он является наименьшим. Аналогичным образом определяются *максимальный* и *наибольший* элементы.

Множество всех подмножеств некоторого множества имеет в качестве наименьшего и наибольшего элементов два несобственных подмножества — $\varnothing$ и само множество. Множество натуральных чисел, упорядоченное отношением делимости, имеет в качестве наименьшего элемента единицу и не имеет наибольшего. Множество разбиений имеет в качестве наибольшего элемента само разбиваемое множество, а в качестве наименьшего — его разбиение на одноэлементные подмножества.

Будем говорить, что b непосредственно следует за a или, что b *покрывает* a, если $a < b$ и между a и b невозможно вставить ни одного элемента c такого, что $a < c < b$. В этом случае будем писать $a \prec b$. Если в конечном упорядоченном множестве $a < b$, но b не покрывает a, то между a и b всегда можно вставить элементы $c_1, \ldots, c_m$ так, чтобы было $a \prec c_1 \prec \ldots \prec c_m \prec b$. Для этого достаточно просто вставлять между a и b элементы до тех пор, пока это возможно. По отношению «$\prec$» легко восстанавливается отношение порядка «$\le$», а именно, $a \le b$ в том и только в том случае, если существует цепочка $a \prec c_1 \prec \ldots \prec c_m \prec b$.

Конечные упорядоченные множества с небольшим числом элементов удобно наглядно представлять с помощью диаграмм Хассе. На ней элементы представляются точками, а непосредственно следующие друг за другом элементы соединяются отрезками, причём конец отрезка, соответствующий большему элементу, располагается выше. Рассмотренные выше система из упорядоченных по включению подмножеств и система упорядоченных по делимости натуральных чисел могут быть представлена диаграммой Хассе как показано на рис. 18.

Диаграммы Хассе ➡

Рассмотрим теперь другой важнейший тип бинарных отношений.

Бинарное отношение на множестве A, являющееся одновременно рефлексивным, симметричным и транзитивным, называется отношением эквивалентности.

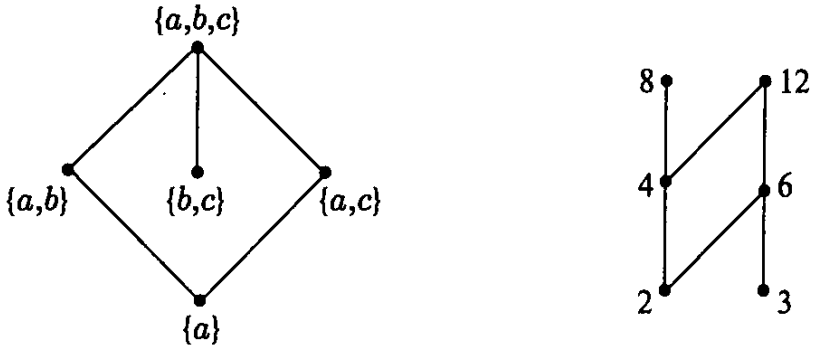


Рис. 18

Отношение эквивалентности

Отношение эквивалентности между элементами a и b обычно записывают как $a \sim b$. Для каждого элемента $a \in A$ множество $K(a)$ всех элементов, эквивалентных a, называется *классом эквивалентности элемента* a. При этом $a \in K(a)$ ввиду рефлексивности отношения эквивалентности.

Любые два класса эквивалентности либо не пересекаются, либо совпадают. В самом деле, пусть $K(a) \cap K(b) \neq \varnothing$ и $c \in K(a) \cap K(b)$. Тогда, если $d \in K(a)$, то $d \sim a \sim c \sim b$ и $d \in K(b)$, откуда $K(a) \subseteq K(b)$. Аналогично доказывается, что $K(b) \subseteq K(a)$, поэтому $K(a) = K(b)$.

Фактормножество

Таким образом, каждое отношение эквивалентности разбивает множество A на непересекающиеся подмножества так, что два элемента лежат в одном классе в том и только в том случае, если они находятся в отношении эквивалентности. Обратно, с любым разбиением множества A на непересекающиеся подмножества можно связать соответствующее отношение эквивалентности. Поэтому подмножества любого разбиения называют *классами*. Множество же классов эквивалентности называется *фактормножеством* множества A по отношению к данному отношению эквивалентности.

Если в качестве множества A взять множество студентов университета, а в качестве бинарного отношения на этом множестве рассмотреть принадлежность двух студентов к одной и той же студенческой группе, то это отношение будет, очевидно, отношением эквивалентности, в котором множество студентов, принадлежащих к одной и той же группе, образует класс эквивалентности. Множество студенческих групп и будет в данном случае фактормножеством.

Отношение параллельности между двумя прямыми на плоскости является, очевидно, симметричным и транзитивным. Каждую прямую можно также считать параллельной самой себе. Так что это есть отношение эквивалентности, в котором множество всех прямых, параллельных данной, образует класс эквивалентности.

Сравнение по модулю ➡

В арифметике важную роль играет отношение *сравнения* на множестве целых чисел $Z = \{0, \pm 1, \pm 2, \pm 3, \ldots\}$. Выберем произвольное натуральное число n и определим на Z отношение R_n следующим образом: $z_1 R_n z_2$, если $z_1 - z_2$ делится на n. Легко проверить, что R_n является отношением эквивалентности. В самом деле, рефлексивность и симметричность проверяются тривиально. Проверим свойство транзитивности. Пусть $z_1 - z_2 = m_1 n$ и $z_2 - z_3 = m_2 n$. Тогда, складывая два равенства, получаем $z_1 - z_3 = (m_1 + m_2) n$, что и доказывает транзитивность. Отношение R_n называется *отношением сравнения по модулю* n. Для него в теории чисел традиционно используется обозначение $z_1 \equiv z_2 \pmod n$, или $z_1 = z_2 \pmod n$. Отношение сравнения играет важную роль во всей дискретной математике.

Любая пара целых чисел сравнима по модулю 1 и всё множество Z при $n = 1$ образует один класс эквивалентности. При $n = 2$ множество целых чисел разбивается на два класса эквивалентности: множество чётных чисел и множество нечётных чисел. В общем случае множество Z разбивается на n классов эквивалентности $K(0), K(1), \ldots, K(n-1)$, характеризуемых остатками от деления на n. Класс $K(r)$ — это множество целых чисел z, представимых в виде $z = r + mn$, где $m \in Z$, что может быть записано как $K(r) = \{r + nZ\}$. Фактормножество $\{K(0), K(1), \ldots, K(n-1)\}$ называются *множеством классов вычетов по модулю* n.

Функции. Если каждому элементу множества A поставлен в соответствие единственный элемент множества B, то говорят, что задано *отображение* из A в B или *функция* $f: A \to B$. Множество A называется областью определения функции f, а B — множеством её значений. Если $b = f(a)$, то элемент b называется *образом* элемента a, а элемент a — *прообразом* элемента b. Множество, состоящее из всех элементов $f(a)$, $a \in A$, называется *образом* множества A при отображении f и обозначается $f(A)$. Ясно, что $f(A) \subseteq B$.

Функцию $y = f(x)$ можно рассматривать и как бинарное отношение R между образом и прообразом (x, y) специального типа, выделяемого условием однозначности: если $(x_1, y_1) \in R$, $(x_2, y_2) \in R$ и $x_1 = x_2$, то $y_1 = y_2$.

Виды функций ➡

Функция $f: A \to B$ называется:

– *инъекцией*, если из $f(a_1) = f(a_2)$ следует $a_1 = a_2$, т. е. никакие два различных элемента множества A не отображаются в один и тот же

элемент множества B, другими словами, каждый элемент множества B имеет не более одного прообраза;

- *сюръекцией*, если для каждого $b \in B$ существует такое $a \in A$, что $b = f(a)$, т. е. множество A отображается на всё множество B, другими словами, каждый элемент множества B имеет хотя бы один прообраз;
- *биекцией*, если f является инъекцией и сюръекцией одновременно, т. е. f устанавливает взаимно однозначное соответствие между множествами A и B.

Эти заимствованные из французского языка термины твёрдо вошли в современную математическую лексику, и их следует запомнить.

Обратная функция

Если функция $f: A \to B$ является биекцией, то для неё существует обратная функция $f^{-1}: B \to A$. Это схематически представлено на рис. 19.

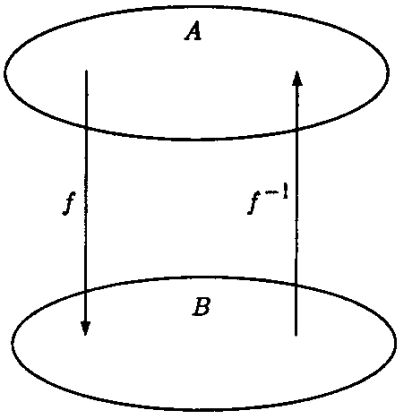


Рис. 19

В математическом анализе для непрерывной функции $f: R \to R$ необходимым и достаточным условием биективности является выполнение одного из двух условий: либо из $x_1 < x_2$ следует $f(x_1) < f(x_2)$, либо из $x_1 < x_2$ следует $f(x_1) > f(x_2)$, т. е. f строго возрастает или строго убывает. Примером такой функции является линейная функция $f: y = ax + b$, если $a \neq 0$. Обратной функцией является $f^{-1}: x = (y - b)/a$. В качестве другого примера можно привести экспоненту $f: y = e^x$, отображающую множество действительных чисел во множество положительных действительных чисел. Обратной к экспоненте является логарифмическая функция $f^{-1}: x = \ln y$.

Для функции, не являющейся биекцией, определить обратную можно, превратив её в биекцию сужением области определения, не изменяющим её образа. В качестве простейшего примера рассмотрим функцию x^2, отображающую множество всех действительных чисел во множество неотрицательных действительных чисел. Определённая таким образом функция сюръективна, но не инъективна и поэтому биекцией не является. Пары одинаковых по модулю, но противоположных по знаку чисел отображаются в одно и то же число, что не позволяет однозначно определить обратную фупкцию. Сузим область определения этой фупкции, заменив множество всех действительных чисел множеством неотрицательных действительных

чисел. Теперь функция x^2, отображающая множество неотрицательных чисел во множество неотрицательных чисел, становится биекцией, и это позволяет определить обратную функцию $\sqrt{x}$, которая называется арифметическим квадратным корнем из x.

Другими известными примерами прямой и обратной функции, получаемых подобным образом, являются $\sin x$ и $\arcsin x$, $\operatorname{tg} x$ и $\operatorname{arctg} x$ и т. д.

Свойство монотонности в самом общем виде формулируется для функций следующим образом. Если на каждом из множеств A и B определено отношение порядка, то функция $f: A \to B$ называется *монотонной*, если из $a_1 \le a_2$ следует $f(a_1) \le f(a_2)$.

Композиция функций ➡

Пусть $f: A \to B$, $g: B \to C$. Тогда можно определить отображение $f \circ g: A \to C$ как $c = g(f(a))$, состоящее в последовательном выполнении сначала отображения f, а затем отображения g, которое называют *композицией* или *суперпозицией* отображений f и g. В курсе математического анализа подобную функцию, однако, традиционно называют сложной функцией. Примерами композиций являются функции $\sin^2 x$, $e^{\sin x}$ и т. д.

Операция композиции отображений ассоциативна. Пусть $f: A \to B$, $g: B \to C$, $h: C \to D$ (рис. 20). Тогда $(f \circ g) \circ h = f \circ (g \circ h)$.

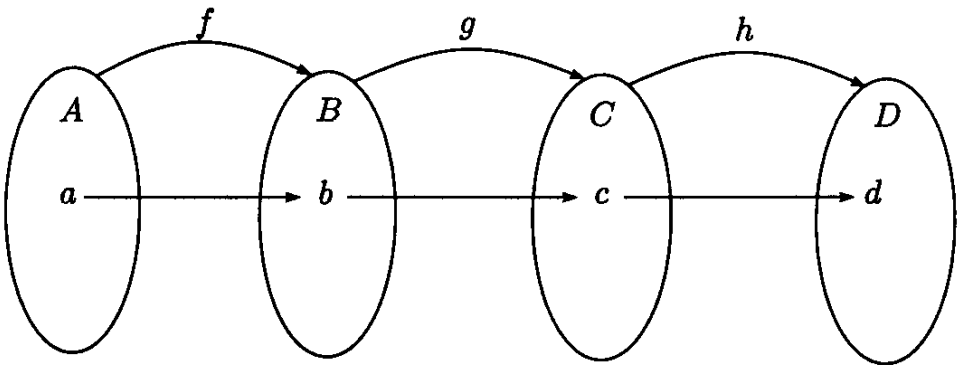


Рис. 20

Сохраняя при записи композиции отображений порядок выполнения отображений слева направо, отображаемый элемент множества удобнее писать слева от символа отображения. Пусть $(a)f = b$, $(b)g = c$, $(c)h = d$. Тогда $(a)(f \circ g) = c$ и $(a)((f \circ g) \circ h) = d$. Аналогично $(b)(g \circ h) = d$ и $(a)(f \circ (g \circ h)) = d$, что и доказывает ассоциативность отображений. Из закона ассоциативности вытекает, что любая расстановка скобок в последовательности отображений приводит к одному и тому же результату.

Композиция прямого и обратного отображения всегда даёт тождественное отображение, например $\ln e^x = e^{\ln x} = x$.

Подстановки. В дискретной математике часто рассматриваются функции $f: A \to B$, где A и B являются конечными множествами. Особый интерес для всей математики представляют биекции вида $f: A \to A$, где A — конечное множество. Такие взаимно однозначные отображения конечных множеств на себя называются *подстановками*. Если A является n-элементным множеством, то подстановка называется *подстановкой степени* n. Пусть $A = \{1, 2, \ldots, n\}$. Тогда подстановку можно записать в виде

$$f = \begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ f(1) & f(2) & f(3) & \ldots & f(n) \end{pmatrix},$$

где нижняя строка является некоторой перестановкой элементов множества $\{1, 2, \ldots, n\}$. Эта строка и задаёт подстановку. Поэтому существует $n!$ подстановок степени n. Среди них имеется и тождественная, оставляющая все элементы на месте и обозначаемая буквой e:

$$e = \begin{pmatrix} 1\,2\,3\,4\,5\,6\,7\,8\,9\,10 \\ 1\,2\,3\,4\,5\,6\,7\,8\,9\,10 \end{pmatrix}.$$

Каждая подстановка может быть разложена в *произведение независимых циклов*. Рассмотрим это разложение на примере. Пусть

$$f = \begin{pmatrix} 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10 \\ 2\ 4\ 6\ 1\ 5\ 7\ 8\ 3\ 10\ 9 \end{pmatrix}.$$

← Разложение на циклы

Тогда $1 \to 2 \to 4 \to 1$, и мы получили первый независимый цикл длины 3, который принято записывать как $(1\ 2\ 4)$. Далее имеем цикл длины 4: $3 \to 6 \to 7 \to 8 \to 3$, который запишется как $(3\ 6\ 7\ 8)$. Элемент 5 остаётся на месте, что может быть записано как (5), а элементы 9 и 10 переставляются местами, т. е. образуют цикл длины 2: $(9\ 10)$. Циклы длины 2 называют *транспозицией*.

Четыре полученных цикла попарно не имеют общих элементов и поэтому называются независимыми. Теперь вся подстановка может быть записана в виде произведения независимых циклов

$$f = (1\ 2\ 4)(3\ 6\ 7\ 8)(5)(9\ 10).$$

Под произведением здесь понимается последовательное выполнение четырёх циклических перестановок, порядок выполнения которых безразличен. Разложение на независимые циклы определяется однозначно, но запись каждого цикла определена с точностью до циклического сдвига: $(3\ 6\ 7\ 8) = (6\ 7\ 8\ 3) = (7\ 8\ 3\ 6) = (8\ 3\ 6\ 7)$. Если подстановка состоит из един-

ственного цикла, то она называется *циклической*. Из $n!$ подстановок степени n циклическими являются $(n-1)!$, так как первый элемент должен перейти один из $n-1$ других элементов, а тот, в свою очередь, в один из оставшихся $n-2$, и т. д.

При записи подстановки в виде произведения независимых циклов циклы длины 1 принято опускать и писать

$$f = (1\ 2\ 4)(3\ 6\ 7\ 8)(9\ 10).$$

Являясь биекцией, каждая подстановка имеет обратную. В нашем случае

Обратная подстановка ➡

$$f^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 4 & 1 & 8 & 2 & 5 & 3 & 6 & 7 & 10 & 9 \end{pmatrix}$$

или $f^{-1} = (1\ 4\ 2)(3\ 8\ 7\ 6)(9\ 10)$.

Композиция подстановок приводит также к подстановке. Так, если

$$g = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 5 & 2 & 9 & 3 & 6 & 4 & 7 & 8 & 1 & 10 \end{pmatrix} = (1\ 5\ 6\ 4\ 3\ 9),$$

то в результате композиции f и g возникает циклическая подстановка:

$$f \circ g = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 1 \end{pmatrix} = (1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10).$$

Композицию подстановок обычно называют их умножением и записывают как произведение: $f \circ g = f \cdot g = fg$. Умножение прямой и обратной подстановок даёт тождественную подстановку:

$$f f^{-1} = f^{-1} f = e.$$

Для любой подстановки f имеет место $fe = ef = f$. В общем же случае произведение подстановок некоммутативно. В качестве подтверждающего примера можно привести рассмотренные выше подстановки f и g:

$$gf = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 5 & 4 & 10 & 6 & 7 & 1 & 8 & 3 & 2 & 9 \end{pmatrix} = (1\ 5\ 7\ 8\ 3\ 10\ 9\ 2\ 4\ 6) \neq fg.$$

Сделаем еще несколько наблюдений, связанных с произведением подстановок. Заметим, что для любых подстановок f и g имеет место $(fg)^{-1} = g^{-1}f^{-1}$, так как

$$(fg)(g^{-1}f^{-1}) = f(gg^{-1})f^{-1} = fef^{-1} = ff^{-1} = e.$$

Пусть некоторая подстановка f раскладывается в произведение k независимых циклов, длины которых $l_1, l_2, \ldots, l_k$. Пусть m есть наимень-

шее общее кратное чисел $l_1, l_2, \ldots, l_k$. Будем рассматривать степени подстановки f: $f^1 = f$, $f^2 = ff$, $f^3 = fff$ и т. д. Тогда $f^m = e$ и m будет первой по порядку степенью, порождающей тождественную подстановку. Более того, все степени $f^1, f^2, \ldots, f^{m-1}$ будут различными подстановками, а $f^{m+1} = f^1$, $f^{m+2} = f^2$, $f^{m+3} = f^3$ и т. д. Число m называется *порядком* подстановки f.

◆ Порядок подстановки

В самом деле, для равенства $f^i = e$ необходимо и достаточно, чтобы i было кратным каждому из чисел $l_1, l_2, \ldots, l_k$. Тогда каждый независимый цикл подстановки f «прокрутится» целое число раз, и все переставляемые элементы останутся на месте. Поэтому m будет первой степенью, для которой $f^m = e$. Все степени $f^1, f^2, \ldots, f^{m-1}$ различны, так как, если бы для $i_1 < i_2 < m$ имело место $f^{i_2} = f^{i_1}$, то это означало бы, что $f^{i_1} f^{i_2 - i_1} = f^{i_1}$. Помножая обе части этого равенства слева на $(f^{i_1})^{-1}$, получаем $(f^{i_1})^{-1} f^{i_1} f^{i_2 - i_1} = (f^{i_1})^{-1} f^{i_1}$, что даёт $f^{i_2 - i_1} = e$, а это невозможно, так как $i_2 - i_1 < m$. Наконец, $f^{m+1} = f^m f = ef = f$, $f^{m+2} = f^m f^2 = ef^2 = f^2$ и т. д. Принято полагать $f^0 = e$. Заметим также, что из равенства $f^m = e$ следует, что $f^{-1} = f^{m-1}$.

Взяв в качестве примера подстановку $f = (123)(45)$, имеем: $f^2 = (132)$, $f^3 = (45)$, $f^4 = (123)$, $f^5 = (132)(45)$, $f^6 = e$, $f^7 = f$ и т. д., $f^{-1} = f^5$.

Группы. Взглянем теперь на множество подстановок с определённой на нём операцией произведения подстановок с более общей точки зрения. Отметим следующие свойства этой алгебраической системы.

1. Произведение двух подстановок $ab = c$ снова является некоторой подстановкой (замкнутость множества подстановок относительно операции произведения).
2. Произведение подстановок ассоциативно: $(ab)c = a(bc)$.
3. Существует тождественная подстановка e, которая, будучи умноженной слева или справа на произвольную подстановку, не изменяет этой подстановки: $ae = ea = a$.
4. Для каждой подстановки a существует обратная ей постановка a^{-1}, которая в произведении с исходной независимо от порядка сомножителей даёт тождественную подстановку: $aa^{-1} = a^{-1}a = e$.

Симметрическая группа ➡

Алгебраическая система, обладающая свойствами 1–4, называется *группой*. Таким образом, множество всех $n!$ подстановок на множестве из n символов с определённой на нём операцией умножения подстановок является группой. Эта группа называется *симметрической группой степени* n и обозначается S_n.

Число элементов в группе называется *порядком группы*. Порядок симметрической группы S_n равен, таким образом, $n!$.

Подгруппы ➡

Некоторое подмножество элементов группы может само образовывать группу. В этом случае оно называется *подгруппой* исходной группы. В частности, подгруппами в S_n являются множества подстановок, порождённые всевозможными степенями произвольной фиксированной подстановки. Подгруппу можно также определить, выделив её элементы с помощью некоторого характеристического свойства, сохраняющегося при операции группового умножения.

Симметрические группы S_n обладают множеством разнообразных подгрупп, многие из которых играют в математике важную роль. Обширный класс таких подгрупп можно получить как группы симметрий определённых объектов, физических или математических, т. е. множество преобразований, совмещающих объект с самим собой. Например, функция $f(x,y,z)=x^2yz+xy^2z+z^2$ не мепяется при транспозиции, меняющей местами переменные x и y, а функция $f(x,y,z)=x^2yz+xy^2z+xyz^2$ не изменяется при любой подстановке переменных x, y, z (такие функции называются симметрическими). Группа симметрий первой функции есть S_2, а второй — S_3. Любой объект воспринимается тем симметричнее, чем обширнее его группа симметрий.

Группы подстановок имеют конечное число элементов, и поэтому называются *конечными группами* или *группами конечного порядка*. Рассматривая биективные преобразования в себя бесконечных множеств, можно получать группы бесконечного порядка.

В качестве примера рассмотрим биективные отображения множества действительных чисел R в себя, задаваемые линейными функциями $f:\ y=ax+b$, где $a\neq 0$. Если $g: y=cx+d$, $c\neq 0$, то $fg:\ y=c(ax+b)+$ $+d=(ac)x+(cb+d)$. Тождественное отображение $e:\ y=x$, обратное отображение $f^{-1}:\ y=x/a-b/a$. Таким образом, эти отображения образуют группу, которая имеет бесконечный порядок. В этой группе можно выделить подгруппу, сохраняющую расстояния между точками на действительной прямой. Элементы этой подгруппы имеют вид $y=x+b$ и $y=-x+b$,

т. е. являются сдвигами и отражениями относительно начала координат с последующим сдвигом.

Подобным же образом можно рассматривать биективные линейные преобразования плоскости или пространства. Преобразования, сохраняющие расстояния между точками, называются *движениями* соответственно прямой, плоскости или пространства. Они образуют подгруппу, называемую группой движений.

⬅ Движения

Движения плоскости намного разнообразнее движений прямой. Сюда входят параллельные переносы, повороты около центра, отражения относительно прямой, а также композиции этих движений. При этом параллельные переносы и повороты образуют подгруппу так называемых *собственных движений* плоскости. Движения, включающие отражение, называются *несобственными*. Собственное движения можно получить как результат непрерывного движения плоскости. Несобственное движения таким образом получить невозможно, так как оно изменяет *ориентацию* — обход произвольного замкнутого контура по часовой стрелке становится после преобразования обходом против часовой стрелки.

Отражения относительно прямой или плоскости называют также зеркальными отражениями. Дважды выполненные зеркальные отражения сохраняют ориентацию, т. е. являются собственными движениями — поворотом или параллельным переносом.

> Алиса ломала себе голову над этими строчками, как вдруг её осенило:
> — Ну конечно, — воскликнула она, — это же Зазеркальная Книга! Если я поднесу её к Зеркалу, я смогу её прочитать.
>
> *Льюис Кэрролл «Алиса в Зазеркалье»*

Связь между группой движений и конечными группами подстановок выявляется при рассмотрении *симметрий* плоских или пространственных фигур. *Симметрией* фигуры называется движение, переводящее эту фигуру в себя. Множество симметрий некоторой фигуры образует группу, являющуюся подгруппой группы движений.

⬅ Симметрии

Если рассматривать симметрии правильного n-угольника на плоскости, то его можно совместить с самим собой n поворотами относительно центра, включая тождественный, на углы $2\pi k/n$, $k = 0, 1, 2, \ldots, (n-1)$, а также отражениями относительно n осей симметрии. В случае чётного n осями симметрии являются $n/2$ прямых, соединяющих середины противоположных сторон и $n/2$ прямых, соединяющих пары противоположных вершин, а в случае нечётного — n прямых, соединяющих вершины с серединами противолежащих сторон. Взятые вместе, n поворотов и n отражений образуют группу симметрий правильного n-угольника, которая имеет порядок $2n$ и называется группой диэдра D_n или диэдральной группой.

⬅ Группа диэдра

Повороты образуют в группе диэдра подгруппу порядка n. Все элементы этой группы могут быть получены последовательным выполнением поворота на угол $2\pi/n$. Если поворот на угол $2\pi k/n$ обозначить через a, то поворот на угол $2\pi k/n$ выразится как a^k. При этом $a^n = e$.

Циклические группы ➡

Группы, порождённые степенями одного элемента, называются *циклическими группами*. Группа поворотов является циклической группой порядка n и обозначается C_n. Каждое отражение порождает циклическую группу C_2. Поэтому говорят, что ось вращения является осью симметрии n-го порядка, а оси отражений — осями симметрий второго порядка. На рис. 21 представлены случаи $n = 5$ и $n = 6$.

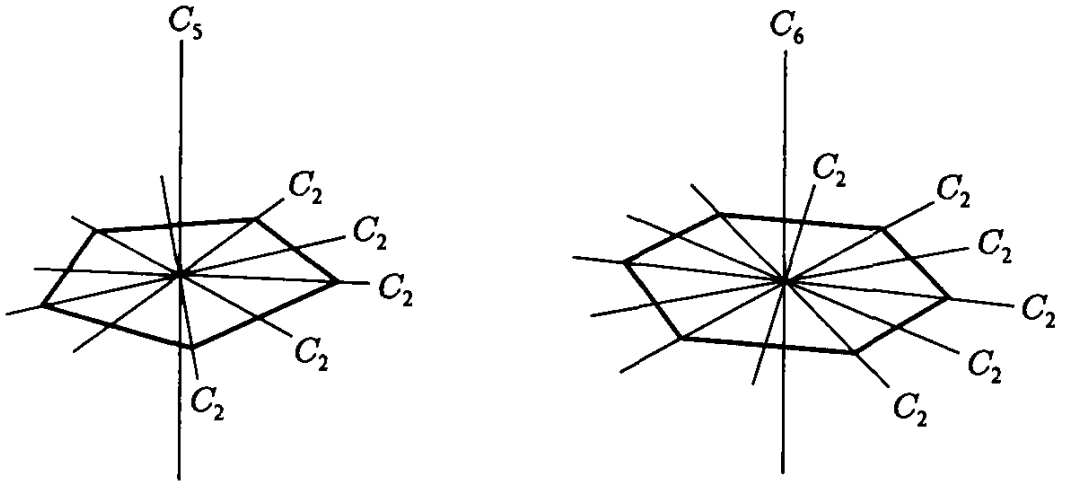


Рис. 21

Пронумеруем вершины n-угольника числами $1, 2, \ldots, n$ по часовой стрелке. На множестве вершин группа диэдра будет действовать как некоторая группа подстановок. Поворот на угол $2\pi/n$ станет циклической перестановкой $(1\ 2\ \ldots, n)$, степени которой порождают циклическую группу C_n. Вся же диэдральная группа перестановок D_n порождается вращением $(1\ 2\ \ldots\ n)$ и любым из отражений, например $\begin{pmatrix} 1\ 2\ 3\ \ldots\ n-1\ n \\ 1\ n\ n-1\ \ldots\ 3\ 2 \end{pmatrix}$. Она является подгруппой симметрической группы S_n.

Симметрии квадрата ➡

В качестве примера рассмотрим подробнее группу самосовмещений квадрата D_4, состоящую из четырёх поворотов на углы 0°, 90°, 180° и 270° и четырёх симметрий относительно осей s_1, s_2, s_3, s_4, которые считаем неподвижными (рис. 22).

Обозначим вращение по часовой стрелке на угол 90° через c. Выберем этот поворот в качестве образующего группы вращений и запишем

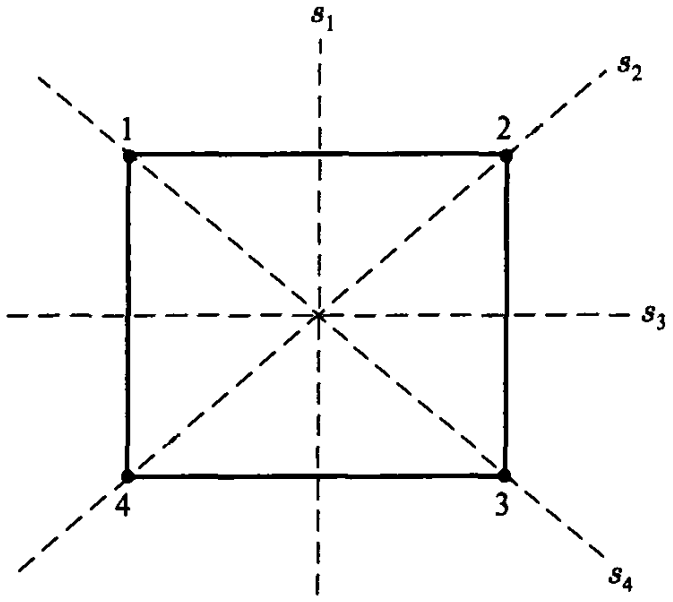


Рис. 22

вращения на 180° и 270°, соответственно, как $cc = c^2$ и $ccc = c^3$. Циклическая группа вращений C_4 состоит, таким образом, из тождественного вращения e на 0° и вращений c, c^2, c^3. Симметрии относительно осей s_1, s_2, s_3, s_4 будем обозначать теми же буквами. Имеем: $c = (1\,2\,3\,4)$, $c^2 = (13)(24)$, $c^3 = (1\,4\,3\,2)$, $s_1 = (12)(34)$, $s_2 = (13)$, $s_3 = (14)(23)$, $s_4 = (24)$. Операция группового умножения может быть задана следующей таблицей (табл. 1), где в клетке, находящейся в строке, помеченной слева элементом g_1, и в столбце, помеченном сверху элементом g_2, стоит элемент $g_1 g_2$. Такие таблицы называют также таблицами Кэли в честь английского математика Артура Кэли (1821–1895).

← Таблица Кэли

·	e	c	c^2	c^3	s_1	s_2	s_3	s_4
e	e	c	c^2	c^3	s_1	s_2	s_3	s_4
c	c	c^2	c^3	e	s_4	s_1	s_2	s_3
c^2	c^2	c^3	e	c	s_3	s_4	s_1	s_2
c^3	c^3	e	c	c^2	s_2	s_3	s_4	s_1
s_1	s_1	s_2	s_3	s_4	e	c	c^2	c^3
s_2	s_2	s_3	s_4	s_1	c^3	e	c	c^2
s_3	s_3	s_4	s_1	s_2	c^2	c^3	e	c
s_4	s_4	s_1	s_2	s_3	c	c^2	c^3	e

Таблица 1

Группа D_4 обладает интересной структурой подгрупп. Во-первых, это *единичная* подгруппа, состоящая из единственного элемента e. Далее, имеется пять циклических подгрупп второго порядка $C_2^0 = \{e, c^2\}$, $C_2^1 = \{e, s_1\}$, $C_2^2 = \{e, s_2\}$, $C_2^3 = \{e, s_3\}$, $C_2^4 = \{e, s_4\}$, а также три подгруппы четвёртого порядка, одна из которых циклическая $C_4 = \{e, c, c^2, c^3\}$, а две другие: $G_1 = \{e, c^2, s_1, s_3\}$, $G_2 = \{e, c^2, s_2, s_4\}$. Структура их взаимных вложений описывается следующей диаграммой Хассе (рис. 23):

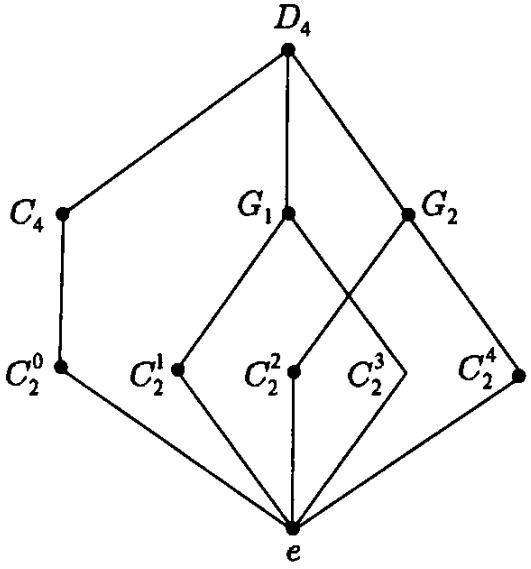


Рис. 23

Заметим, что в группе восьмого порядка D_4 имеются подгруппы первого, второго и четвёртого порядков, но нет подгрупп третьего, пятого, шестого или седьмого порядков. Это обстоятельство не является случайным. Порядок подгруппы, как мы увидим в дальнейшем, всегда является делителем порядка группы.

Истоки теории групп ➡

Теория групп первоначально возникла как теория групп подстановок при изучении проблемы разрешимости алгебраических уравнений в радикалах. С глубокой древности известна изучаемая в школе формула для решения квадратного уравнения. Итальянскими математиками в эпоху Возрождения были найдены формулы для решения уравнения третьей и четвёртой степени. Однако попытки дальнейшего продвижения в этом направлении не увенчались успехом. Формул для решения уравнений пятой и более высоких степеней, несмотря на значительные усилия многих математиков, найти не удавалось.

Поворотным пунктом в решении этой проблемы стали исследования крупнейшего французского математика и механика второй половины XVIII века Жозефа Луи Лагранжа (1736–1813). Рассматривая подстановки корней уравнения, Лагранж сделал первые шаги в создании теории групп и впервые предположил отсутствие общих формул для решения уравнений степени выше четвёртой. Эти исследования были продолжены итальянским математиком Паоло Руффини (1765–1822) и норвежским математиком Нильсом Абелем (1802–1829), который, изучая группы подстановок, впервые строго доказал отсутствие общих формул для решения алгебраических уравнений степени выше 4-й.

Окончательное решение проблемы разрешимости в радикалах алгебраических уравнений удалось найти гениальному французскому матема-

тику Эвариcту Галуа (1811–1832), который в возрасте двадцати лет, занимаясь исследованиями в этой области, заложил основы теории групп и дал мощный толчок развитию всей современной алгебры. Им, в частности, был введён и сам термин «группа». Судьба Галуа сложилась трагически. Страстный республиканец, за выступления против Луи-Филиппа он попадает в тюрьму, а вскоре после выхода из неё погибает на дуэли. Стояла ли за этим любовная интрига или это была замаскированная под неё политическая расправа — осталось тайной, покрытой мраком. Зная, что погибнет, в ночь накануне дуэли Галуа пишет письмо другу и спешно правит свои математические записи, содержащие идеи, которые математический мир смог понять и оценить лишь много лет спустя после его гибели.

✓ *Галуа*

За прошедшее с тех пор время понятие группы постепенно выкристаллизовалось в одно из самых фундаментальных сначала в алгебре, а затем и во всей математике. Тесно связанное с симметрией, оно возникает в самых различных её разделах. Чтобы обладать такой общностью, понятие группы определяется аксиоматически. Поэтому, познакомившись с группами на неформальном уровне, перейдём к точным определениям.

◆ Определение абстрактной группы

Говорят, что на множестве M определена некоторая бинарная алгебраическая операция $*$, если определено отображение $M^2 \to M$, т. е. каждой упорядоченной паре (a, b) элементов из M поставлен в соответствие некоторый однозначно определённый элемент $c \in M$, что записывается как $a*b=c$. Множество M вместе с заданной на нём бинарной операцией $*$, удовлетворяющей определённым условиям, и образует группу. В рассматриваемых ранее группах преобразований элементами множества M были биективные отображения некоторого множества на себя, а операцией $*$ — их композиция.

Множество M с определённой на нём бинарной операцией $*$ называется группой $G=<M; *>$, если бинарная операция $*$ удовлетворяет следующим трём свойствам:

1) ассоциативна: $(ab)c=a(bc)$ для любых $a,b,c \in M$;

2) существует элемент $e \in M$, называемый единичным или нейтральным элементом группы, такой, что $ea=ae=a$ для любого $a \in M$;

3) для любого $a \in M$ существует элемент a^{-1}, называемый обратным к a, такой, что $a^{-1}*a=a*a^{-1}=e$.

Определив группу аксиоматически, можно, основываясь только на этом определении, получить ряд общих результатов. Докажем, прежде

всего, единственность нейтрального элемента e. Пусть наряду с e существует нейтральный элемент e'. Тогда имеем

$$e' = e * e' = e,$$

где левое равенство написано на основании того, что e является нейтральным элементом, а правое — что нейтральным элементом является e'. Используя транзитивность равенства, заключаем, что $e' = e$.

Далее, уравнение $a * x = b$ всегда имеет в группе единственное решение. В самом деле, умножая слева обе части этого равенства на a^{-1} и пользуясь ассоциативным законом, получаем:

$$a^{-1} * (a * x) = a^{-1} * b \Rightarrow (a^{-1} * a) * x = a^{-1} * b \Rightarrow x = a^{-1} * b.$$

Аналогично, решением уравнения $y * a = b$ является $y = b * a^{-1}$.

Отсюда следует, в частности, единственность обратных элементов, а также свойство таблицы Кэли быть латинским квадратом. Заметим также, что, как и в группе подстановок, в любой группе $(a * b)^{-1} = b^{-1} * a^{-1}$.

Коммутативные группы ➡

Группами являются также ряд числовых множеств со сложением или умножением в качестве групповой операции. Умножение и сложение коммутативны. Если групповая операция коммутативна, т. е. для любых двух элементов группы выполнено $a * b = b * a$, то группа G называется *коммутативной* или *абелевой*, в честь Нильса Абеля.

Множество $R \setminus \{0\}$ отличных от нуля действительных чисел с обычным арифметическим умножением в качестве групповой операции является абелевой группой $< R \setminus \{0\}; \times >$. Нейтральным (единичным) элементом этой группы является число 1. Обратным к элементу $a \in R \setminus \{0\}$ является $a^{-1} = 1/a$.

Подгруппой группы $< R \setminus \{0\}; \times >$ является группа $< R^+; \times >$ всех положительных действительных чисел. Другой подгруппой группы $< R \setminus \{0\}; \times >$ является группа $< Q \setminus \{0\}; \times >$ отличных от нуля рациональных чисел, подгруппой которой в свою очередь является группа $< Q^+; \times >$ положительных рациональных чисел. Вложимость этих групп друг в друга представлена диаграммой Хассе на рис. 24.

Подчеркнём, что для того чтобы некоторое множество $H \subseteq M$ образовывало подгруппу группы $G = < M; * >$ необходимо и достаточно, чтобы выполнялись следующие условия:

1) $e \in H$;
2) если $a, b \in H$, то $a * b \in H$;
3) если $a \in H$, то $a^{-1} \in H$.

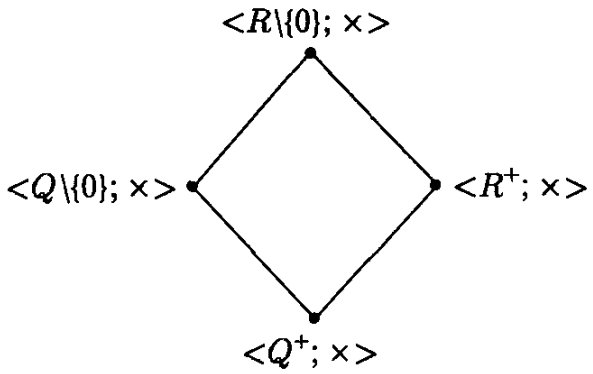


Рис. 24

В дальнейшем, если групповая операция фиксирована, группа и её подгруппы будут отождествляться с множествами их элементов. В этом случае будем говорить просто о группе G и её подгруппе H.

Всё множество R действительных чисел будет также коммутативной группой $< R; + >$, если в качестве групповой операции рассматривать арифметическую операцию сложения. Нейтральным элементом здесь является число 0, а обратным к числу a — противоположное ему число $(-a)$.

Подгруппами группы $< R; + >$ будет группа $< Q; + >$ всех рациональных чисел, а также группа $< Z; + >$ всех целых чисел. Группа $< Z; + >$, в свою очередь, обладает счётным множеством подгрупп. А именно, для каждого натурального n множество $nZ = \ldots, -2n, -n, 0, n, 2n, \ldots$ всех целых чисел, кратных n, с операцией сложения будет подгруппой $< nZ; + >$ этой группы. Вложимость этих числовых подгрупп представлена диаграммой Хассе на рис. 25.

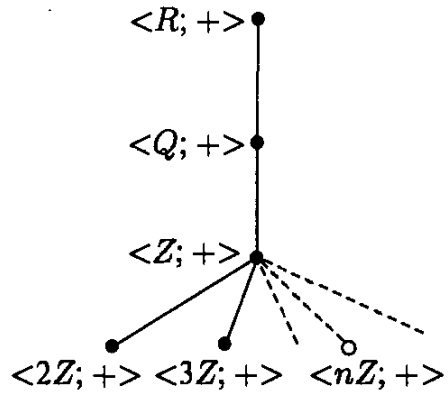


Рис. 25

В векторном пространстве относительно операции сложения векторы образуют коммутативную группу. Абстрактное векторное пространство так и определяется в современной алгебре. Это абелева группа, дополненная операцией умножения элементов группы (векторов) на действительные числа (или их обобщения — элементы произвольного поля).

Далее групповую операцию $a * b$ удобно будет записывать как обычное умножение ab.

Пусть $a, b, \ldots$ — произвольные элементы некоторой группы G. Всевозможные конечные произведения этих элементов и обратных

◂ Образующие

к ним, т. е. произведения вида $a^{-1}a^{-1}bbab^{-1}$, образуют подгруппу группы G. Говорят, что эта подгруппа *порождена* элементами $a, b, \ldots$, которые называют её *образующими*. Так группа диэдра D_n порождается, например, вращением на угол $2\pi/n$ и любым из отражений.

Циклические группы ➡

Группа, порождённая одним элементом, т. е. состоящая из положительных и отрицательных степеней одного элемента, называется *циклической*. Циклические группы — это простейшие по своей алгебраической структуре, но весьма важные группы. Рассмотрим их подробнее.

Циклическая группа является коммутативной группой, так как $a^i a^j = a^j a^i = a^{i+j}$. Далее, для циклической группы существуют две возможности. Либо все степени $a^0, a^1, a^2, a^3, \ldots$ различны, тогда циклическая группа $\ldots, a^{-2}, a^{-1}, a^0, a^1, a^2, \ldots$ бесконечна. В качестве примера можно привести аддитивную группу $< nZ; + >$, бесконечную для любого $n \in N$.

Либо $a^m = e$, и циклическая группа, порождённая элементом a, имеет порядок m и состоит из элементов $e, a^1, a^2, \ldots, a^{m-1}$, здесь $a^{-k} = a^{m-k}$. В этом случае говорят, что *порядок элемента* a равен m. Подобный случай возникал при изучении группы подстановок. Циклическую группу порядка m обозначают C_m.

Пусть C_m — циклическая группа порядка m с образующим элементом a. Возьмём в C_m элемент a^i, где i взаимно просто с m, и рассмотрим последовательность его степеней $(a^i)^1 = a^i$, $(a^i)^2 = a^{2i}$, $(a^i)^3 = a^{3i}, \ldots$. Пусть j — первая степень, для которой $(a^i)^j = a^{ij} = e$. Тогда ij должно быть кратно m, а так как i взаимно просто с m, то это возможно лишь при $j = m$. Поэтому циклическая группа, порождаемая элементом a^i, совпадёт с исходной циклической группой. Это имеет место для тех и только для тех i, которые взаимно просты с m. Отсюда вытекает, что C_m имеет $\varphi(m)$ образующих, где $\varphi(m)$ — фупкция Эйлера.

Подгруппы и факторгруппы. Пусть G — произвольная группа, H — её подгруппа. На множестве элементов группы G рассмотрим отношение R, задаваемое следующим образом: $(a, b) \in R \Leftrightarrow a^{-1}b \in H$. Это отношение рефлексивно, так как $a^{-1}a = e \in H$. Далее, если $a^{-1}b \in H$, то и $(a^{-1}b)^{-1} \in H$, но $(a^{-1}b)^{-1} = b^{-1}(a^{-1})^{-1} = b^{-1}a$. Поэтому из $(a, b) \in R$ следует, что

$(b, a) \in R$, т. е. отношение R симметрично. Наконец, если $a^{-1}b \in H$ и $b^{-1}c \in H$, то $a^{-1}c = (a^{-1}b)(b^{-1}c) \in H$, т. е. отношение R транзитивно. Поэтому отношение R является отношением эквивалентности, и в дальнейшем его будем записывать как $a \sim b$.

Как и каждое отношение эквивалентности, это отношение разбивает множество элементов группы G на классы эквивалентности. Рассмотрим эти классы внимательней. Если $a \sim b$, то $a^{-1}b = h$, где $h \in H$, и $b = ah$. Наоборот, из $b = ah$ следует, что $a \sim b$, т. е. a и b принадлежат одному и тому же классу эквивалентности. Поэтому класс элементов, эквивалентных элементу a, имеет вид aH, т. е. состоит из произведений элемента a на все элементы подгруппы H. Множество $aH = \{ah \mid h \in H\}$ называется *левым смежным классом по подгруппе H, порождаемым элементом a*. При этом $a \in aH$, и смежный класс aH порождается любым элементом $g \in aH$.

Множество элементов группы G распадается, таким образом, на непересекающиеся левые смежные классы по подгруппе H. Это разложение называется *левосторонним разложением группы G по подгруппе H*. При этом одним из смежных классов является сама подгруппа H.

◆ Разложение по подгруппе

Заметим, что рассмотренные ранее классы вычетов $K(0), K(1), \ldots, K(n-1)$ по модулю n являются смежными классами группы $G = < Z; + >$ по подгруппе $H = < nZ; + >$.

Левостороннее разложение группы D_4 по подгруппе $C_4 = \{e, c_1, c_2, c_3\}$ состоит из двух смежных классов: $\{e, c_1, c_2, c_3\}$ и $\{s_1, s_2, s_3, s_4\}$, а по подгруппе $C_2^0 = \{e, c_2\}$ из четырёх смежных классов: $\{e, c_2\}$, $\{c_1, c_3\}$, $\{s_1, s_3\}$ и $\{s_2, s_4\}$.

Каждый смежный класс равномощен подгруппе H. Можно сказать, что любая подгруппа «укладывается» в группе целое число раз. Это приводит к следующему важному результату.

Теорема Лагранжа. В конечной группе порядок любой подгруппы является делителем порядка группы.

Мощность множества смежных классов в разложении группы G по подгруппе H, которая в конечном случае согласно теореме Лагранжа равна $|G|/|H|$, называется *индексом* подгруппы H в группе G.

Делителем порядка конечной группы согласно теореме Лагранжа будет и порядок любого её элемента. Если порядок группы равен n, а порядок некоторого её элемента a равен m, то $a^m = e$ и $a^n = (a^m)^{n/m} = e$. Отметим этот результат.

Следствие 1. Любой элемент конечной группы, возведённый в степень, равную порядку группы, даёт единичный элемент группы.

В качестве ещё одного следствия из теоремы Лагранжа получаем

Следствие 2. Каждая подгруппа простого порядка является циклической и порождается любым своим элементом, отличным от e.

Аналогично левым смежным классам и левостороннему разложению группы G по подгруппе H можно было бы рассматривать *правые смежные классы*, т. е. множества вида Ha, и разбиение группы G на множество правых смежных классов, которое называется *правосторонним разложением группы G по подгруппе H*. Это также приводит к теореме Лагранжа. В общем случае левый и правый смежные классы по подгруппе, порождаемые одним и тем же элементом могут не совпадать. Весьма важен, однако, случай, когда для любого $a \in G$ имеет место $aH = Ha$, т. е. левый и правый смежные классы, порождаемые элементом a, совпадают.

Нормальный делитель ➡

Если для любого элемента $g \in G$ порождаемый им левый смежный класс по подгруппе H совпадает с правым смежным классом $gH = Hg$, то подгруппа H называется нормальной подгруппой группы G или нормальным делителем этой группы.

Любая подгруппа индекса 2 является, очевидно, нормальным делителем, а в коммутативной группе вообще любая подгруппа является нормальным делителем. В других случаях вопрос о нормальности данной подгруппы требует специального исследования. Если обратиться к рассмотренной ранее группе диэдра D_4, то её подгруппы C_4, G_1 и G_2 являются нормальными делителями как подгруппы индекса 2. А из пяти подгрупп индекса 4 нормальной является лишь подгруппа $C_2^0 = \{e, c^2\}$. В самом деле, имеем $s_1\{e, c^2\} = s_1\{e, s_1s_3\} = \{s_1, s_3\}$ и $\{e, c^2\}s_1 = \{e, s_3s_1\}s_1 = \{s_1, s_3\}$. Аналогично, $s_2\{e, c^2\} = s_2\{e, s_2s_4\} = \{s_2, s_4\}$ и $\{e, c^2\}s_2 = \{e, s_4s_2\}s_2 = \{s_2, s_4\}$. А также, $c\{e, c^2\} = \{c, c^3\}$ и $\{e, c^2\}c = \{c, c^3\}$.

Понятие нормального делителя является чрезвычайно важным в алгебре, так как множество смежных классов по нормальной подгруппе само образует группу, если в качестве произведения двух смежных классов aH и bH рассмотреть множество $(aH)(bH)$ всевозможных произведений элементов из одного смежного класса на элементы другого. Это множество в случае нормального делителя является смежным классом:

$$(aH)(bH) = (aH)(Hb) = a(HH)b = aHb = a(Hb) = a(bH) = (ab)H.$$

Можно также сказать, что произведение двух классов смежности — это тот смежный класс, в который попадает произведение двух произвольных элементов этих классов. Единичным элементом в определённой таким образом группе классов смежности по нормальному делителю H является сама подгруппа H. Подгруппа, образованная таким образом из смежных классов, называется *факторгруппой* группы G по нормальному делителю H и обозначается G/H.

◆ Факторгруппа

В качестве примера рассмотрим факторгруппу D_4/C_2^0. Обозначив её элементы как $C_2^0 = \{e, c^2\} = e$, $\{c, c^3\} = a$, $\{s_1, s_3\} = b$, $\{s_2, s_4\} = c$, получаем следующую таблицу умножения в этой группе (табл. 2)

Как видно из табл. 2 группа D_4/C_2^0 оказалась коммутативной группой, в которой каждый элемент является обратным к самому себе.

·	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

Таблица 2

Взглянем теперь на способ определения понятия факторгруппы с несколько иных, более общих позиций. На множестве с операцией группового умножения было выбрано отношение эквивалентности, которое оказалось согласованным с операцией умножения в том смысле, что из $a \sim a'$ и $b \sim b'$ следует $ab \sim a'b'$. Именно в результате такой согласованности и оказалось, что класс эквивалентности, в который попадает произведение, зависит только от классов, из которых выбираются сомножители, но не от конкретных представителей этих классов. Это и позволило перенести на фактормножество групповую операцию, образовав факторгруппу. Отношение эквивалентности, согласованное подобным образом с алгебраическими операциями на множестве, называют *конгруэнцией*, что в переводе с латинского и означает *согласие, соответствие*.

Кольца и поля. Наряду с группами основными рассматриваемыми в современной алгебре структурами являются также *кольца* и *поля*. Эти новые названия не должны смущать впервые знакомящегося с ними читателя. Вскоре он обнаружит в них хорошо знакомые ему со школы объекты и может почувствовать себя в положении мольеровского господина Журдена, воскликнувшего:

> Честное слово, я и не подозревал, что вот уже более сорока лет говорю прозой.
>
> *Мольер «Мещанин во дворянстве»*

◆ Кольцо целых чисел

Правда, читатель может спросить, зачем нужны новые названия для уже известных объектов. Ответ состонт в том, что это позволяет выделять существенные свойства различных алгебраических структур, единообразно их рассматривать и строить новые структуры с заданными свойствами.

Непустое множество M с двумя операциями «+» и «×» называется кольцом $<M; +, \times>$, если:

1) $< M; + >$ является коммутативной группой;

2) умножение удовлетворяет левому и правому законам дистрибутивности относительно сложения:

$$a(b+c)=ab+bc \text{ и } (b+c)a=ba+ca.$$

Группа $< M; + >$ называется аддитивной группой кольца, нейтральный элемент которой обозначается символом 0. Если операция умножения «×» ассоциативна, обладает единичным элементом, коммутативна, то кольцо называется соответственно ассоциативным, кольцом с единицей, коммутативным.

Классическим примером кольца является кольцо $<Z; +,\times>$ — множество целых чисел Z с операциями сложения и умножения. Это кольцо ассоциативное, с единицей, коммутативное. Как и во всяком кольце кроме операций сложения и умножения, в нём всегда выполнима также операция вычитания: $a-b=a+(-b)$. А вот операция деления выполнима не всегда. Единственными обратимыми элементами этого кольца являются 1 и -1.

Кольцо многочленов ➡

Другим важным примером кольца является множество всех многочленов с действительными коэффициентами с операциями сложения и умножения. Напомним, что многочленом степени n с действительными коэффициентами называется выражение вида

$$a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0,$$

где

$$a_i \in R,\ i=0,1,2,\ldots,n,\ a_n \neq 0.$$

Причём в алгебре многочлен обычно понимается не как функция переменного x, а просто как формальное выражение, в записи которого используется буква x. Сложение, умножение и вычитание многочленов осуществляется по обычным школьным правилам. Это кольцо также ассоциативное, с единицей и коммутативное. Как и в кольце целых чисел, деление здесь выполнимо не всегда, но всегда выполнимо деление с остатком. Для любых двух многочленов $f(x)$ и $g(x)$ существуют однозначно определённые многочлены $q(x)$ и $r(x)$ такие, что

$$f(x)=q(x)g(x)+r(x),$$

где степень $r(x)$ меньше степени $g(x)$ или же $r(x)=0$, когда деление выполнимо нацело. Выполнять деление многочленов с остатком можно с помощью «деления уголком». Например, если требуется разделнть $2x^2+3x-1$ на $x+1$, то деление уголком даёт

$$\begin{array}{l} \_2x^2+3x-1 \quad \underline{|x+1} \\ \underline{2x^2+2x} \qquad 2x+1 \\ \_x-1 \\ \underline{x+1} \\ -2\,(\textit{остаток}), \end{array}$$

что позволяет записать

$$2x^2+3x-1=(2x+1)(x+1)-2\,.$$

В качестве остатка здесь оказалось число, т. е. многочлен нулевой степени.

Если некоторый многочлен $f(x)$ мы делим на $x-c$, то в остатке всегда будет некоторое число

$$f(x)=(x-c)q(x)+r\,.$$

Если в это равенство подставить $x=c$, то получим $f(c)=r$. В частности, если c — корень многочлена $f(x)$, т. е. $f(c)=0$, то остаток r оказывается равным нулю. Мы получили, таким образом, теорему Безу, утверждающую, что, если c — корень многочлена $f(x)$, то многочлен делится на $x-c$ нацело. В качестве следствия из теоремы Безу получается известное со школы разложение квадратного трёхчлена на линейные множители:

$$x^2+px+q=(x-x_1)(x-x_2)\,,$$

где x_1 и x_2 — корни квадратного трёхчлена.

Кольцо целых чисел и кольцо многочленов имеют между собой много общего. В обоих кольцах выполнимо деление с остатком, величина которого меньше величины делителя. В кольце целых чисел за величину числа принимается его модуль, а в кольце многочленов — степень многочлена. Роль простых чисел в кольце многочленов играют многочлены, которые не имеют делителей степени выше нулевой. Такие многочлены называются *неприводимыми*. Неприводимым, например, является многочлен x^2+1. В кольце многочленов также справедлива теорема о единственности разложения на простые множители, в качестве которых выступают неприводимые многочлены.

◆ Неприводимые многочлены

В то же время между двумя кольцами имеется существенная разница. В кольце целых чисел существуют сколь угодно большие простые числа. В кольце же многочленов с действительными коэффициентами каждый многочлен, степени большей двух, представим как произведение двух многочленов меньшей степени. Так, например, любой кубический многочлен раскладывается в произведение многочленов первой и второй степеней.

В качестве примера некоммутативного кольца можно привести множество вещественных квадратных матриц порядка n с операциями сложения и умножения. Это ассоциативное кольцо с единицей, в качестве которой выступает единичная матрица, а нулевым элементом здесь является матрица, целиком состоящая из нулей.

Со школьной скамьи известно, что нуль, умноженный на любое число, даёт нуль. Этим же свойством нуль обладает в кольце многочленов и в кольце квадратных матриц. Остаётся ли данное свойство нуля справедливым в любом кольце? Ответ положителен, в любом кольце имеет место $0 \cdot a = a \cdot 0 = 0$.

При доказательстве справедливости этого в произвольном кольце можно опираться лишь на свойства кольца, постулируемые в определении. Имеем $a + 0 = a$. Помножив обе части этого равенства справа на a, получим $a^2 + 0 \cdot a = a^2$. Теперь прибавление к обеим частям равенства $-a^2$ даёт $0 \cdot a = 0$. Аналогично доказывается, что $a \cdot 0 = 0$.

Правило знаков $(-a) \cdot b = -(a \cdot b)$ также справедливо в любом кольце, так как $a \cdot b + (-a) \cdot b = (a + (-a)) \cdot b = 0 \cdot b = 0$. Таким образом, многие из привычных алгебраических преобразований выполнимы в каждом кольце.

Кольцо классов вычетов ➧

Введём теперь важные для дискретной математики кольца из конечного числа элементов. С этой целью покажем, что отношение сравнения по натуральному модулю n, являющееся, как известно, отношением эквивалентности, является конгруэнцией в кольце целых чисел. В самом деле, пусть $z_1' \equiv z_1 (\mathrm{mod}\, n)$, $z_2' \equiv z_2 (\mathrm{mod}\, n)$. Тогда $z_1' = z_1 + k_1 n$, $z_2' = z_2 + k_2 n$ и $z_1' + z_2' = z_1 + z_2 + (k_1 + k_2)n$, $z_1' z_2' = z_1 z_2 + (k_1 z_2 + k_2 z_1 + k_1 k_2 n)n$. То есть, $z_1' + z_2' \equiv z_1 + z_2 (\mathrm{mod}\, n)$ и $z_1' z_2' = z_1 z_2 (\mathrm{mod}\, n)$.

Доказанная конгруэнтность сравнения позволяет стандартным способом определить на множестве классов вычетов $\{K(0), K(1), \ldots, K(n-1)\}$ операции сложения и умножения. Результатом операции над двумя классами является класс, в который попадает результат соответствующей арифметической операции над произвольно выбранными двумя представителями данных классов. Необходимые для определения кольца свойства следуют из соответствующих свойств кольца целых чисел. Нулевым элементом кольца является класс $K(0) = \{nZ\}$, а единичным — класс $K(1) = \{1 + nZ\}$. Определённое таким образом факторкольцо называется *кольцом классов вычетов по модулю* n и обозначается Z/nZ.

Так как любое целое число z единственным образом представимо в виде

$$z = kn + r, \text{ где } k \in Z,\ r \in N \text{ и } 0 \le r < n,$$

причём каждому классу вычетов соответствует своё значение остатка r, то в качестве представителей классов удобно выбрать числа $0,1,2,\ldots,n-1$ и считать их элементами кольца.

Заметим также, что для любого элемента a кольца Z/nZ имеет место соотношение

$$\underbrace{a+a+\ldots+a}_{n\ раз}=0\,.$$

В качестве примера рассмотрим кольцо $Z/6Z$. Элемент $Z_r=r+6Z$ этого кольца будем записывать просто как r. Сложение и умножение в этом кольце задаются следующими таблицами (табл. 3).

+	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

×	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

Таблица 3

Если привычным геометрическим образом кольца целых чисел являются целые точки на прямой (рис. 26а), то естественным геометрическим образом кольца вычетов по модулю n являются точки на окружности (рис. 26б).

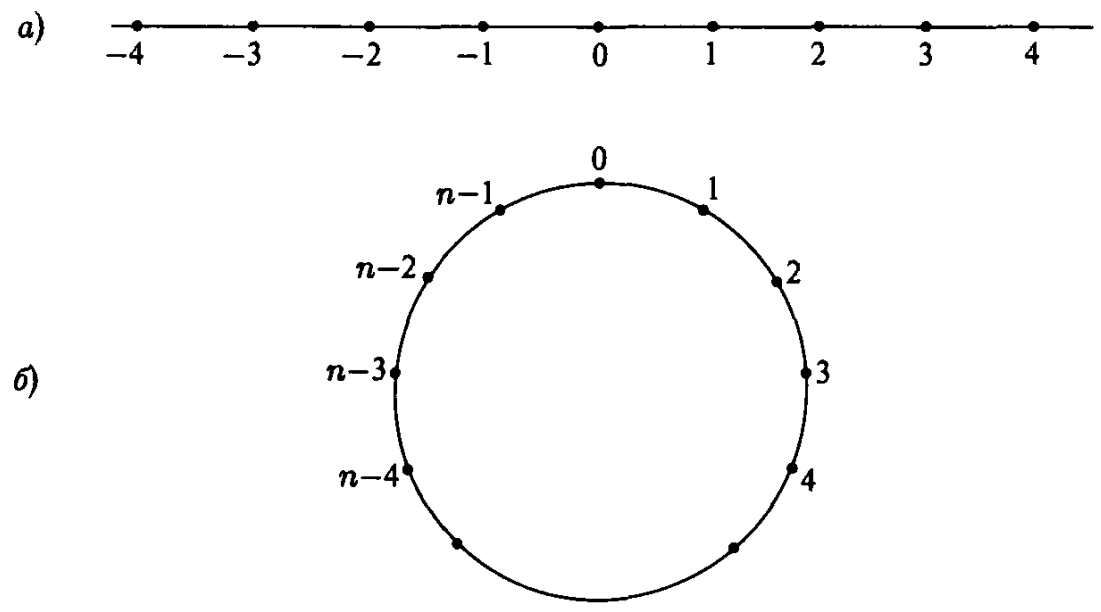


Рис. 26

Множество всех действительных чисел с операциями сложения и умножения является ассоциативным и коммутативным кольцом с единицей. Однако оно обладает ещё одним важным свойством — любой ненулевой элемент этого поля обратим, что позволяет в этом кольце наряду с вычитанием выполнять также и деление на любой ненулевой элемент: $a:b=\frac{a}{b}=ab^{-1}$.

Ассоциативное и коммутативное кольцо с единицей <M; +, ×> называется полем, если каждый ненулевой элемент кольца обратим, т. е. < M\{0}; × > является коммутативной группой. Группа <M\{0}, ×> называется мультипликативной группой поля, нейтральный элемент в ней обозначается 1 и называется единицей.

В полях выполнимы все четыре привычных арифметических действия. Вложенными друг в друга числовыми полями являются поле рациональных чисел, поле действительных чисел и поле комплексных чисел: $Q \subset R \subset C$. Это важнейшие числовые поля. Теперь, говоря о многочленах, мы будем рассматривать не только многочлены, коэффициенты которых являются действительными числами — многочлены над полем действительных чисел, но и многочлены над любыми полями, например, полем рациональных или комплексных чисел, каждый раз указывая, над каким полем рассматриваются многочлены. Алгоритм деления с остатком и единственность разложения на неприводимые множители справедливы для многочленов над любыми полями, но приводимость или неприводимость многочлена зависит от того, над каким полем он рассматривается. Например, многочлен x^2-2 неприводим над полем рациональных чисел, но приводим над полем действительных чисел: $x^2-2=(x-\sqrt{2})(x+\sqrt{2})$.

Числовые поля ➡

Для дискретной математики, однако, особенно важными являются поля, состоящие из конечного множества элементов. Рассмотрим таблицу умножения в кольце $Z/5Z$ (табл. 4).

Конечные поля ➡

×	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Таблица 4

Заметим, что для любого ненулевого элемента этого кольца в соответствующей ему строке таблицы присутствуют все элементы этого кольца, в том числе и единица. А это значит, что все ненулевые элементы этого кольца обратимы и оно является полем. Имеем $1^{-1}=1$, $2^{-1}=3$, $3^{-1}=2$, $4^{-1}=4$. Поэтому, как и в каждом поле, здесь возможно деление на любой ненулевой элемент, например $2:3=4$, так как $2:3=2\cdot 3^{-1}=$ $=2\cdot 2=4$.

Покажем, что подобным свойством обладает кольцо вычетов Z/pZ по любому простому модулю p и, значит, все такие кольца являются полями. Пусть p простое. Докажем, что все ненулевые элементы $\{1,2,\ldots,p-1\}$ кольца Z/pZ обратимы относительно операции умножения. С этой целью возьмём произвольный ненулевой элемент r, где $1 \le r \le p-1$, и рассмотрим его произведения на все ненулевые элементы: r, $2r$, $\ldots$, $(p-1)r$. Все полученные таким образом $p-1$ элементов будут ненулевыми, так как ir при $i < p$ и $r < p$ не может быть кратным простому числу p. Все они будут также попарно различны, так как, если бы среди них оказалась пара одинаковых ir и jr, то число $(j-i)r$ было бы кратным p, что невозможно, так как $j-i < p$ и $r < p$, а p — число простое. Поэтому $(r, 2r, \ldots, (p-1)r)$ является перестановкой элементов множества $\{1,2,\ldots,p-1\}$ и в этой перестановке обязан присутствовать единичный элемент.

Тем самым доказано, что в кольце Z/pZ элементы $\{1,2,\ldots,p-1\}$ образуют коммутативную группу по умножению и это кольцо является полем. Условие простоты модуля n является не только достаточным, но и необходимым для того, чтобы кольцо Z/nZ было полем. В самом деле, если $n = ij$, то произведение двух соответствующих элементов кольца обращается в нуль, т. е. выходит за пределы множества ненулевых элементов. Поэтому ненулевые элементы кольца не образуют группу по умножению.

Поля Z/pZ называют *конечными полями*, а также *полями Галуа* в честь Эвариста Галуа и обозначают GF(p) (Galois Field — поле Галуа). Простейшим полем Галуа является поле GF(2), состоящее из двух элементов 0 и 1. Сложение и умножение в нём приведены в табл. 5.

+	0	1
0	0	1
1	1	0

×	0	1
0	0	0
1	0	1

Таблица 5

Как и в каждой группе, в мультипликативной группе поля Z/pZ, имеющей порядок $p-1$, любой элемент i удовлетворяет соотношению $i^{p-1} = 1$. Возвращаясь к определению поля Z/pZ и вспоминая, что его элементы являются классами вычетов, получаем следующий важный результат.

> **Малая теорема Ферма.** Пусть p простое. Тогда для каждого натурального i, не кратного p, имеет место $i^{p-1} \equiv 1 \pmod{p}$.

Эта теорема, установленная великим французским учёным Пьером Ферма, играет важную роль в теории чисел.

Расширения полей. Если некоторое подмножество элементов поля само является полем относительно тех же операций сложения и умножения, то

оно называется *подполем* большего поля, а большее поле — *надполем* или *расширением* меньшего. Так поле действительных чисел является расширением поля рациональных, а поле комплексных чисел — расширением поля действительных. Если имеется неприводимый над некоторым полем многочлен $f(x)$ степени не меньшей двух, то это поле можно расширить, рассмотрев на множестве всех многочленов над данным полем классы вычетов по $f(x)$. Два многочлена при этом попадают в один класс, если их разность делится на $f(x)$ нацело. Эти классы образуют кольцо, которое в случае неприводимого многочлена является полем. Положение здесь в точности такое же, как и в случае вычетов в кольце целых чисел. В качестве представителей классов естественно брать многочлены наименьшей степени в этом классе, т. е. остатки от деления на $f(x)$. При этом классы вычетов, представленные многочленами нулевой степени, образуют подполе, совпадающее с исходным полем.

В качестве примера возьмём неприводимый над полем рациональных чисел многочлен x^2-2. В качестве представителей классов берём многочлены минимальной степени в этом классе, т. е. остатки от деления на x^2-2. Эти остатки являются многочленами первой степени над полем рациональных чисел, и каждый элемент получающегося поля классов вычетов может быть записан как $ax+b$, где a и b — рациональные числа.

Алгебраические действия над элементами поля выполняются по обычным правилам. Сложение вообще не вызывает затруднения, например,

$$(3x+2)+(4x+3)=7x+5.$$

При умножении же происходит повышение степени многочлена, поэтому оно требует выполнения операции приведения к многочлену наименьшей степени в данном классе вычетов. Имеем

$$(3x+2)(4x+3)=12x^2+17x+6.$$

Но

$$12x^2+17x+6=12(x^2-2)+17x+30\ .$$

Поэтому

$$6+17x+12x^2\equiv 30+17x \mod(x^2+1)$$

и

$$(2+3x)(3+4x)=30+17x\ .$$

Сам многочлен x^2-2 попадает при этом в нулевой класс. Таким образом, в нашем поле $x^2-2=0$, т. е. элемент x является корнем многочлена x^2-2, откуда $x^2=2$. Это позволяет считать x таким добавленным к множеству рациональных чисел числом, квадрат которого равен 2, и обозначить его как $\sqrt{2}$. Теперь элементы поля будут записываться как

$a\sqrt{2}+b$, где $a,b \in Q$, что позволяет весьма наглядно выполнять действия над ними. Например, если требуется выполнить деление, то

$$\frac{2+3\sqrt{2}}{3+4\sqrt{2}} = \frac{(2+3\sqrt{2})(3-4\sqrt{2})}{(3+4\sqrt{2})(3-4\sqrt{2})} = \frac{-18+\sqrt{2}}{3^2-\left(4\sqrt{2}\right)^2} = \frac{-18+\sqrt{2}}{-23} = \frac{18}{23} - \frac{1}{23}\sqrt{2}.$$

Если у элемента поля $a\sqrt{2}+b$ коэффициент $a=0$, то этот элемент может считаться обычным рациональным числом b. Действия с такими элементами выполняются как действия с рациональными числами и не выводят за множество рациональных чисел. Таким образом, рассматриваемое поле есть расширение поля рациональных чисел, полученное присоединением к нему корня многочлена x^2-2 — иррационального числа $\sqrt{2}$. Построенное поле является надполем поля рациональных чисел и подполем поля действительных чисел.

Аналогичным образом можно взять неприводимый над полем действительных чисел многочлен x^2+1 и считать классы вычетов по нему элементами нового поля, беря в качестве их представителей остатки $ax+b$, где $a,b \in R$. Элемент x в этом поле является корнем уравнения $x^2+1=0$ и удовлетворяет соотношению $x^2=-1$. Этот элемент принято обозначать буквой i, называя его мнимой единицей и записывая элементы поля в виде $ai+b$. Действия в этом поле выполняются согласно правилам

$$(ai+b)+(ci+d) = (a+c)i+(b+d);$$
$$(ai+b)(ci+d) = (ad+bc)i+(bd-ac).$$

Построенное поле является полем *комплексных чисел*. Расширить подобным образом это поле уже невозможно, так как в нём нет неприводимых многочленов степени выше первой.

Но вернёмся к конечным полям. Кроме полей GF(p), образованных классами вычетов по простому модулю, для всех натуральных n и простых p существуют также поля GF(p^n), состоящие из p^n элементов. Других конечных полей нет. Поля GF(p^n) также можно получить как классы вычетов по неприводимому над полем GF(p) многочлену

$$a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0,$$

где

$$a_i \in GF(P),\ i = 0,1,2,\ldots,n,\ a_n \neq 0.$$

Оказывается, что неприводимые многочлены степени n над полем GF(p) существуют для любого натурального n.

Как и в случае расширений числовых полей, элементами поля GF(p^n) являются классы многочленов, определяемые остатками от деления на

выбранный неприводимый многочлен. Такими остатками являются все многочлены над GF(p) степени, не большей чем $n-1$. Эти остатки можно взять в качестве представителей классов. Так как многочлен степени $n-1$ имеет n коэффициентов, каждый из которых может принимать p значений, то всего имеется p^n остатков, и поле $GF(p^n)$ состоит из p^n элементов.

При сложении многочленов коэффициенты при одинаковых степенях складываются. Поэтому, если произвольный элемент a поля Галуа $GF(p^n)$ сложить сам собой p раз, то все коэффициенты многочлена, представляющего сумму, обратятся в нуль и сумма будет равна нулю:

$$\underbrace{a+a+\ldots+a}_{p\ раз}=0.$$

В качестве примера построим поле $GF(2^2)$. Неприводимым многочленом второй степени над GF(2) является многочлен x^2+x+1. В его неприводимости легко убедиться следующим образом. Если бы он раскладывался на множители, то каждый из них был бы многочленом первой степени и имел в качестве корня один из элементов поля GF(2), которых всего два: 0 и 1. Тогда этот же элемент был бы корнем и исходного многочлена. Но, как показывает простая проверка, ни одно из этих значений не обращает данный многочлен в нуль.

Элементами нашего поля, таким образом, являются: 0, 1, x, $x+1$. При сложении элементов x и $x+1$ получаем

$$x+(x+1)=(x+x)+1=0+1=1,$$

а умножение этих же элементов даёт

$$x\times(x+1)=x^2+x=(x^2+x+1)+1\equiv 1 \bmod (x^2+x+1).$$

Полностью таблицы сложения и умножения представлены в табл. 6.

+	0	1	x	$x+1$
0	0	1	x	$x+1$
1	1	0	$x+1$	x
x	x	$x+1$	0	1
$x+1$	$x+1$	x	1	0

×	0	1	x	$x+1$
0	0	0	0	0
1	0	1	x	$x+1$
x	0	x	$x+1$	1
$x+1$	0	$x+1$	1	x

Таблица 6

Изоморфизм. *Изоморфизм* является чрезвычайно важным понятием для всей современной математики. Возникшее первоначально в алгебре, оно постепенно проникло и в другие её разделы. Название образовано из греческих слов «изо» — равный, одинаковый, подобный, и «морфе» — вид, форма, образ. Оно описывает схожесть двух алгебраических систем с задан-

ными в них операциями или отношениями. Системы называются *изоморфными*, если между двумя множествами, на которых заданы операции или отношения, можно установить такое взаимно однозначное соответствие, которое сохраняет эти операции и отношения. Рассмотрим это понятие детальнее.

Пусть на равномощных множествах A и B заданы соответственно бинарные отношения R_1 и R_2, также имеющие одинаковую мощность. Тогда системы $< A, R_1 >$ и $< B, R_2 >$ называются *изоморфными*, если существует биекция $f: A \to B$, переводящая одно отношение в другое, т. е. $(f(a_1), f(a_2)) \in R_2 \Leftrightarrow (a_1, a_2) \in R_1$. Изоморфизм между системами $< A, R_1 >$ и $< B, R_2 >$ обозначается как $< A, R_1 > \cong < B, R_2 >$. Отношение изоморфизма является, разумеется, отношением эквивалентности.

Пусть, например, $A=\{a_1, a_2, a_3\}$, $R_1=\{(a_1,a_2), (a_1,a_3)\}$, $B=\{b_1,b_2,b_3\}$, $R_2=\{(b_2,b_3), (b_2,b_1)\}$. Тогда $< A; R_1 > \cong < B; R_2 >$, изоморфизм между двумя системами устанавливается биекцией $f: \begin{pmatrix} a_1 & a_2 & a_3 \\ b_2 & b_3 & b_2 \end{pmatrix}$. Схематически две изоморфные системы отношений вместе с устанавливающей изоморфизм биекцией показаны на рис. 27, где отношения в каждой из систем указаны стрелками. Заметим, что биекция $f: \begin{pmatrix} a_1 & a_2 & a_3 \\ b_2 & b_1 & b_3 \end{pmatrix}$ также задаёт изоморфизм.

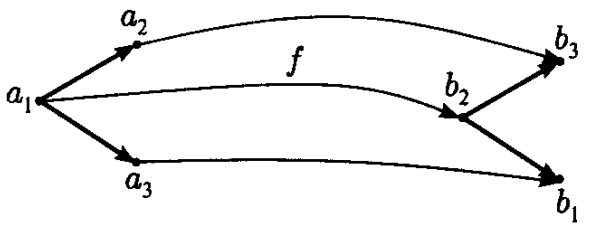


Рис. 27

Аналогичным образом понятие изоморфизма определяется для множеств с операциями. Пусть $\varphi: A^2 \to A$ и $\psi: B^2 \to B$ бинарные операции во множествах A и B. Тогда $< A;\varphi > \cong < B;\psi >$, если существует биекция $f: A \to B$, согласованная с операциями во множествах: $f(\varphi(a_1,a_2)) = \psi(f(a_1), f(a_2))$ (образ результата операции есть результат операции над образами). Множество B с операцией ψ является как бы зеркальным отражением множества A с операцией φ.

С алгебраической точки зрения, фокусирующей внимание на свойствах самих отношений и операций, изоморфные системы неразличимы. Классическим примером изоморфизма алгебраических систем является изоморфизм множества действительных чисел с операцией сложения $< R; + >$ и

множества положительных действительных чисел с операцией умножения $< R^+; \times >$. Биекцией $f: R \to R^+$, устанавливающей этот изоморфизм, является функция $f(x) = e^x$, так как $e^{x_1+x_2} = e^{x_1} \cdot e^{x_2}$. Обратная биекция $f^{-1}: R^+ \to R$ — это функция $\ln x$: $\ln(x_1 \cdot x_2) = \ln x_1 + \ln x_2$ (рис. 28).

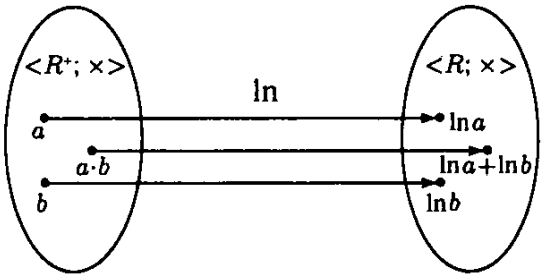


Рис. 28

Данный изоморфизм позволяет сводить умножение к сложению. Для нахождения произведения двух чисел x и y достаточно найти их логарифмы, сложить их, а затем, произведя обратную биекцию, вычислить экспоненту от суммы: $xy = e^{(\ln x + \ln y)}$. На этом основано выполнение умножения с помощью логарифмической линейки, широко использовавшейся в инженерных расчётах до появления компьютеров. Две обычные линейки с равномерными шкалами позволяют производить сложение двух длин, приставляя к концу одной длины начало другой. Если же линейки размечены так, что расстояние от начала линейки до отметки числа равно логарифму этого числа, то, сумме двух длин будет соответствовать отметка произведения чисел.

✓*Джон Непер*

Честь открытия логарифмов принадлежит шотландскому математику барону Джону Неперу (1550–1617). После завершения образования в Эдинбурге и путешествия по Европе он в возрасте двадцати одного года навсегда поселился в семейном поместье близ Эдинбурга, где занялся богословием и математикой. Богословские сочинения Непера, направленные против католицизма и Папы, в эпоху Реформации принесли ему даже большую известность, чем математические открытия, но потом были забыты. Вызванная же к жизни его открытием логарифмическая линейка, состоящая из двух одинаковых шкал, одна из которых перемещается вдоль другой, появилась в середине XVII века и просущеетвовала до второй половины двадцатого в качестве основного вычислительного средства в инженерных расчётах.

Изоморфизм групп ➡

Рассмотрим теперь один вопрос, относящийся к изоморфизму групп в самой общей постановке. До того, как дать формальное определение группы как множества с бинарной операцией, обладающей определёнными свойствами, в качестве примеров групп рассматривались группы подстановок на конечных и бесконечных множествах с суперпозицией подстановок в качестве групповой операцией. Оказывается, что эти примеры и являются самыми общими примерами абстрактных групп, и любая группа, как было подмечено Кэли, изоморфна некоторой группе подстановок.

Теорема Кэли. Каждая группа изоморфна некоторой группе подстановок.

Для доказательства этого каждому элементу $g \in G$ сопоставим подстановку Gg на множестве элементов группы G. Тогда в силу ассоциативности произведению элементов будет соответствовать суперпозиция двух подстановок

$$G(g_1 g_2) = (G g_1) g_2 .$$

Таким образом, рассматривая группы подстановок, мы не проигрываем в общности.

Теперь можно коснуться и интересного вопроса об изоморфизме *абстрактных групп* конечного порядка. Термином «абстрактных» подчёркивается, что не имеет значения, каким конкретным способом задаётся та или иная группа, задана она как группа подстановок конечного множества или каким-то другим образом. Две группы изоморфны, если существует биекция, сохраняющая групповую операцию, т. е. переводящая таблицу Кэли одной группы в таблицу Кэли другой. Нейтральные элементы при этом, разумеется, отображаются друг в друга. Класс изоморфных групп и принято называть *абстрактной группой.*

Заметим, прежде всего, что все циклические группы одинакового порядка попарно между собой изоморфны. Изоморфизм между двумя группами устанавливается отождествлением одинаковых степеней образующих. Если группы G_1 и G_2 являются циклическими группами одинакового порядка n и элемент α является образующим в циклической группе G_1, а элемент β — в группе G_2, то биекция $\alpha^i \leftrightarrow \beta^i$, $i = 0,1,2,\ldots,n-1$ является, очевидно, изоморфизмом и $G_1 \cong G_2$. А так как любая группа простого порядка является циклической, то все группы простого порядка изоморфны между собой, т. е. существует одна абстрактная группа заданного простого порядка, а именно циклическая группа. Таким образом, существует по одной абстрактной группе порядков 2, 3, 5, 7 и т. д.

Рассмотрим теперь группы четвёртого порядка. Если в группе существует элемент четвёртого порядка, то этот элемент порождает группу, и, следовательно, группа циклическая. Если же в группе нет элементов четвёртого порядка, то, как вытекает из теоремы Лагранжа, квадрат любого элемента есть единичный элемент. Произведение любых двух неединичных элементов этой группы может быть равно только третьему неединичному элементу. Этими условиями таблица Кэли задаётся уже однозначно (табл. 2).

Абстрактную группу, задаваемую таблицей 2 и возникавшую ранее как факторгруппа D_4 / C_2^0, в теории групп называют *четверной группой Клейна* по имени немецкого математика Феликса Клейна (1849–1925), первым подметившего связи между геометрией и теорией групп.

Конкретными реализациями четверной группы Клейна являются рассмотренные ранее подгруппы G_1 и G_2 группы симметрий квадрата D_4 (рис. 23). Каждая из этих подгрупп состояла из тождественного преобра-

зования и трёх симметрий — одной центральной и двух осевых. Фигурой, группа симметрий которой изоморфна четверной группе Клейна, является ромб. Группа же симметрий параллелограмма изоморфна группе C_2, так как помимо тождественного преобразования она содержит лишь центральную симметрию.

Таким образом, существуют две различные группы четвёртого порядка — циклическая и четверная группа Клейна. Можно показать также, что существуют 2 группы шестого порядка — циклическая и изоморфная S_3, 5 групп восьмого порядка и по 2 группы девятого и десятого порядков.

Что же касается других абстрактных алгебраических систем, то отметим, что поле Галуа $GF(p^n)$ для каждого натурального n и простого p единственно с точностью до изоморфизма.

Графы. Простейший способ наглядного представления конечного множества состоит в изображении его элементов точками на плоскости. Пусть $A = \{a_1, a_2, a_3\}$. Тогда множество A можно представить в наглядной форме как показано на рис. 29.

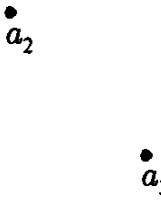


Рис. 29

Однако такой способ может оказаться не слишком удобен, если рассматриваются подмножества множества A.

Пусть снова $A = \{a_1, a_2, a_3\}$. Тогда множество A имеет 8 подмножеств: $\varnothing$, $\{a_1\}, \{a_2\}, \{a_3\}, \{a_1, a_2\}$, $\{a_1, a_3\}, \{a_2, a_3\}, \{a_1, a_2, a_3\}$ с характеристическими векторами соответственно: $(0,0,0), (1,0,0), (0,1,0), (0,0,1)$, $(1,1,0), (1,0,1), (0,1,1), (1,1,1)$. Множество подмножеств в этом случае можно визуализировать, считая подмножества вершинами единичного куба, помещенного в положительном октанте, так что характеристические векторы являются координатами вершин (рис. 30).

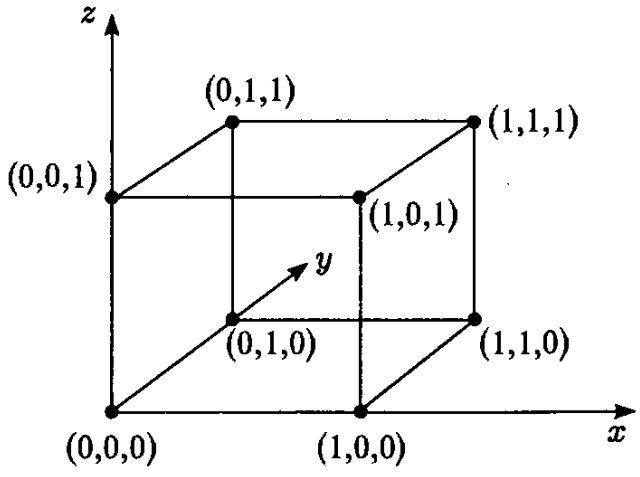


Рис. 30

Заметим, что, начав с 3-элементного множества, лишённого какой бы то ни было структуры, т. е. без связей или отношений между его элементами, мы получили 8-элементное множество с интересной структурой. Ребра куба соединяют те подмножества, которые отличаются друг от друга добавлением или выбрасыванием одного элемента, т. е. выражают отношение близости между подмножествами.

Убрав оси координат и оставив только точки, обозначающие подмножества, и ребра, выражающие отношение близости между подмножествами, получим *граф* — фигуру, состоящую из точек (*вершин графа*) и соединяющих их линий (*рёбер графа*) (рис. 31).

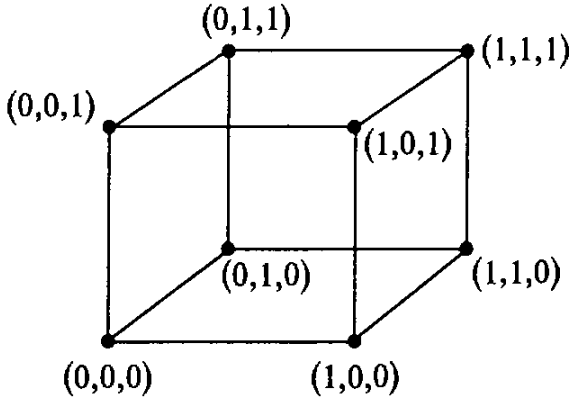


Рис. 31

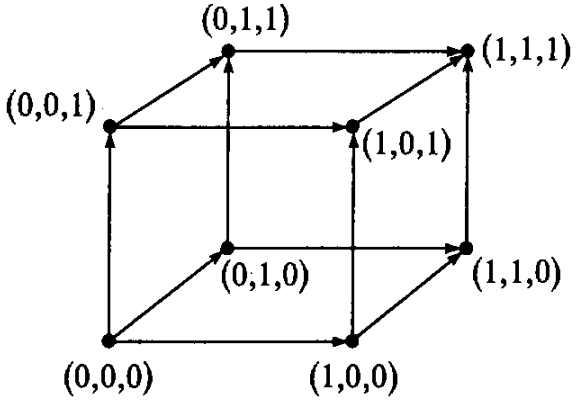


Рис. 32

Чтобы указать, какое из двух соседних подмножеств больше, поставим на рёбрах стрелки, идущие от меньшего подмножества к большему (рис. 32).

Получился *ориентированный граф* (*орграф*), состоящий из вершин и рёбер со стрелками, которые принято называть *дугами*. Говорят, что стрелки задают ориентацию рёбер и превращают их в дуги. Если ребро — это неупорядоченная пара вершин, то дуга — упорядоченная. Подобно диаграмме Хассе данный ориентированный граф определяет на подмножествах отношение порядка по включению. Ориентированный граф задает данный частичный порядок в том смысле, что одно подмножество содержит другое тогда и только тогда, когда из меньшего подмножества можно перейти в большее, двигаясь по дугам графа в направлении стрелок. Вообще, произвольный орграф задаёт на множестве своих вершин отношение порядка в том и только в том случае, если в орграфе отсутствуют замкнутые пути (*контуры*).

◆ Орграфы

Графы и орграфы являются весьма важными объектами в дискретной математике. Чтобы проиллюстрпровать их значение, рассмотрим ещё пример. Пусть имеется некоторое множество людей, которых условно обозначим буквами $\{a,b,c,d,e\}$. Пусть среди них есть следующие пары знакомых между собой: $\{a,b\},\{a,c\},\{b,e\},\{c,e\},\{c,d\}$. Тогда отношение знакомства можно задать следующим графом (рис. 33).

◆ Граф знакомств

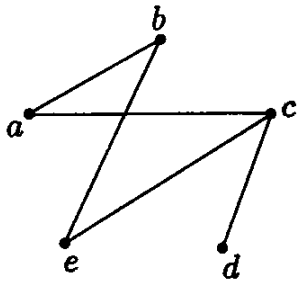


Рис. 33

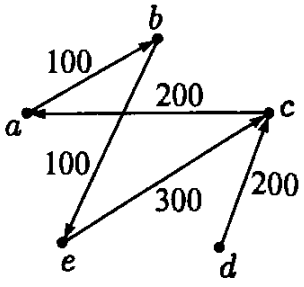


Рис. 34

Отношение знакомства является симметричным отношением, так как если a знаком с b, то b знаком с a. Поэтому такое отношение задаётся обычным (неориентированным) графом.

Пусть теперь известно, что a должен b 100 рублей, c должен a 200 рублей, b должен e 100 рублей e должен c 300 рублей, d должен c 200 рублей. Тогда эту информацию можно задать с помощью орграфа, приписав каждой дуге величину долга (рис. 34).

Граф или орграф, рёбрам или дугам которого приписаны положительные числа (веса), называется *взвешенным*. Таким образом, информация о должниках была задана с помощью взвешенного орграфа.

Взвешенный орграф долговых обязательств ➡

При графическом задании информации наглядно видно, что система долговых обязательств может быть упрощена с помощью взаимозачётов, если в графе существует замкнутый путь (*контур*). Найдя на этом пути дугу минимального веса и вычтя этот вес из весов всех дуг пути, получим ориентированный граф с меньшим числом дуг, если уберём дуги нулевого веса. При этом каждое лицо, входящее в замкнутый путь, прощает часть своего долга предшествующему лицу и получает прощение такой же суммы долга от последующего лица.

И прости нам долги наши, как и мы прощаем должникам своим.

Евангелие от Матфея, 6.12

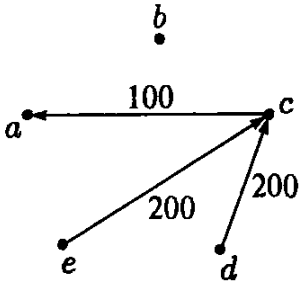


Рис. 35

В нашем примере контуром, по которому может быть произведено взаимное погашение долгов, является контур (a,b,e,c,a). Вычтя из весов всех его дуг 100, получим граф с упрощённой системой долговых обязательств (рис. 35).

Систему из выделенных подмножеств конечного множества также можно представить в виде графа, множество вершин которого разбито на два подмножества (доли): подмножество вершин, соответствующих элементам исходного множества, и подмножество вершин, соответствующих его подмножествам. Если элемент принадлежит подмножеству, то соответствующие вершины соединены ребром. Таким образом, рёбра могут соединять только вершины из разных долей. Такие графы называются *двудольными*. Пусть, например, в исходном множестве $A=\{a_1,a_2,a_3,a_4,a_5\}$ выделены подмножества $A_1=\{a_1,a_2,a_3\}$, $A_2=\{a_1,a_4\}$, $A_3=\{a_2,a_5\}$, $A_4=\{a_5\}$. Тогда двудольный граф, представляющий данную систему подмножеств, выглядит следующим образом (рис. 36).

Двудольные графы ➡

Его называют также графом *инциденций* системы элементы — подмножества.

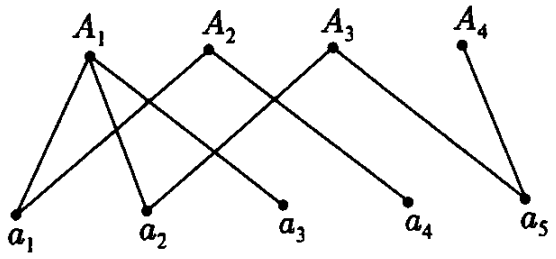


Рис. 36

Чтобы понять, как в практической деятельности могут возникать двудольные графы, представим себе бюро по трудоустройству, в которое обращаются лица, желающие получить работу. Текущая информацию в таком бюро может быть задана с помощью двудольного графа, одну долю которого составляют обратившиеся в бюро лица, а другую — имеющиеся в распоряжении бюро вакантные места. Ребро в таком графе означает, что данное лицо соответствует данному вакантному месту по образованию, полу, возрасту и т. д.

Граф можно также представлять себе и как некоторую коммуникационную сеть, считая, например, вершины городами, а рёбра — соединяющими их дорогами. При этом длины дорог будут весами соответствующих рёбер. Тогда естественным образом возникает задача о нахождении кратчайшего пути из одной вершины взвешенного графа в другую. Разумеется, она может быть решена полным просмотром всех возможных разумных путей. Однако такой просмотр может привести к огромному перебору, неосуществимому даже на компьютере. Рассмотрим в качестве примера следующий граф (рис. 37).

⬅ Маршруты в графах

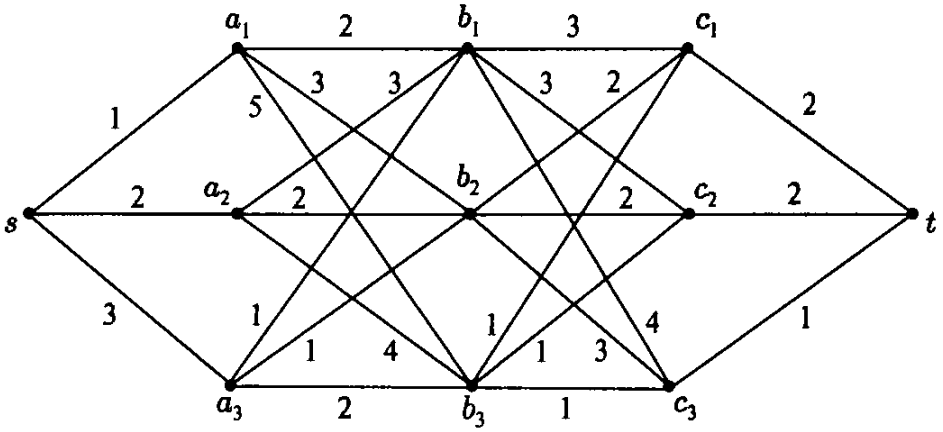


Рис. 37

Кратчайший маршрут из s в t имеет в данном случае длину 7, и читатель сможет легко найти его. Полный перебор, однако, требует просмотра $3\times3\times3 = 27$ вариантов маршрутов:

$$(s,a_1,b_1,c_1,t),\ (s,a_1,b_1,c_2,t),\ (s,a_1,b_1,c_3,t),\ (s,a_1,b_2,c_1,t),$$

$$(s,a_1,b_2,c_2,t),\ \ldots,\ (s,a_3,b_3,c_2,t),\ (s,a_3,b_3,c_3,t).$$

И это для графа всего лишь с 11 вершинами! Число возможных маршрутов стремительно нарастает с ростом числа вершин в графе. Для графа с несколькими сотнями вершин полный перебор при выборе оптимального

маршрута может оказаться не под силу даже мощному компьютеру. Поэтому задача состоит в нахождении такого алгоритма, число операций которого бы не слишком быстро возрастало с ростом числа вершин, например, для некоторого фиксированного небольшого k было бы ограничено величиной n^k от числа вершин графа n. Такой алгоритм действительно существует. Заинтересованный читатель найдёт его на страницах этой книги.

Задача нахождения эффективных алгоритмов или доказательство их отсутствия является важнейшей проблемой современной дискретной математики, полностью не решённой до сих пор. На протяжении книги читатель встретит множество алгоритмов, эффективность которых будет обсуждаться.

Доказательства от противного. Двоичные наборы и графы являются наиболее часто встречающимися объектами в дискретной математике. Что же касается способов логических рассуждений и доказательств, то остановимся на двух из них — доказательстве от противного, как мощном средстве, используемом во всей математике, и методе математической индукции, особенно часто применяемом в дискретной математике.

Доказательство от противного в древности было принято называть доказательством путём приведения к абсурду. Этот метод доказательства фактически уже использовался нами при доказательстве существования бесконечных несчётных множеств и доказательстве того, что мощность 2^A превышает мощность A. В качестве других классических примеров рассмотрим найденные ещё в Древней Греции доказательства бесконечности множества простых чисел и иррациональности числа $\sqrt{2}$.

Бесконечность простых чисел ➡

Напомним, что отличное от единицы натуральное число p называется простым, если оно делится только на 1 и на себя. Первыми простыми числами являются 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, ... Является ли этот ряд бесконечным или простых чисел конечное число? В Началах Евклида имеется замечательное доказательство его бесконечности.

Для того, чтобы доказать, что простых чисел бесконечно много, допустим противное, что простых чисел конечное число. Пусть $P = \{p_1, p_2, \ldots, p_k\}$ — множество всех простых чисел. Тогда все отличные от единицы и не входящие в P натуральные числа являются составными и обязаны иметь в качестве делителей числа из P, так как каждое составное число раскладывается на простые множители.

Образуем число $p = p_1 p_2 \ldots p_k + 1$. Так как это число не делится ни на одно из чисел $p_1, p_2, \ldots, p_k$ и отлично от единицы, то оно должно быть простым и обязано совпадать с одним из чисел $p_1, p_2, \ldots, p_k$, но это невозможно, так как p больше всех чисел из P. Полученное противоречие показывает абсурдность сделанного предположения и доказывает бесконечность ряда простых чисел.

Докажем теперь тем же самым приёмом иррациональность числа $\sqrt{2}$. Предположим противное, что $\sqrt{2}$ является рациональным числом, т. е. представимо в виде несократимой дроби $\sqrt{2} = \frac{p}{q}$. Возводя обе части равенства в квадрат, получаем $2q^2 = p^2$. Откуда следует, что p делится на 2. Пусть $p = 2k$. Тогда $q^2 = 2k^2$. Отсюда вытекает, что q также делится на 2. Но это противоречит сделанному предположению о несократимости дроби $\frac{p}{q}$.

◆ Иррациональность $\sqrt{2}$

Иррациональность $\sqrt{2}$ имела для греческой математики огромное значение, так как показывало несоизмеримость диагонали квадрата с его стороной, нарушая тем самым, по мнению греков, канонический порядок вещей. Несоизмеримость означает отсутствие отрезка, который бы целое число раз укладывался как в стороне квадрата (q раз), так и в его диагонали (p раз). В самом деле, если бы такой отрезок длины a существовал, то по теореме Пифагора было бы $p^2a^2 = 2q^2a^2$, откуда $\sqrt{2} = \frac{p}{q}$. Платон сообщает, что был потрясён, когда в юности впервые узнал об этом.

Математическая индукция. Метод математической индукции более молод. Его зарождение относится к эпохе Возрождения, а первое применение в явном виде встречается у Франческо Мавролико (1494–1575) из Мессины, помимо математики занимавшегося также физикой, метеорологией и историей и сделавшего полный перевод сочинений Архимеда. Систематическое же использование метода полной математической индукции в математике началось после работ Блеза Паскаля.

Этот метод доказательства основан на *принципе математической индукции*, согласно которому утверждение $A(n)$, зависящее от натурального параметра n, верно при всех n, если верно $A(1)$ и если из справедливости $A(n)$ следует справедливость $A(n+1)$. Проверка $A(1)$ называется *базой индукции*, а доказательство справедливости $A(n+1)$ в предположении справедливости $A(n)$ — *индуктивным переходом*. Метод математической индукции часто используется при доказательстве эмпирически подмеченной закономерности.

В качестве простейшего примера применения этого метода возьмём рассмотренную у Мавролико задачу о сумме первых n нечётных чисел.

◆ Сумма нечётных чисел

Имеем: $1 = 1^2$, $1 + 3 = 2^2$, $1 + 3 + 5 = 3^2$, $1 + 3 + 5 + 7 = 4^2$ и т. д. Эти наблюдения подсказывают общее утверждение:

При любом натуральном n сумма первых n нечётных чисел равна n^2.

Докажем это предложение по индукции. Оно верно при $n = 1$. Пусть для некоторого n сумма первых n нечётных натуральных чисел равна n^2. Тогда сумма первых $n+1$ нечётных чисел равна $n^2 + (2n+1) = (n+1)^2$. Этим и доказывается данное утверждение.

Разбиение плоскости прямыми ➡

Рассмотрим другую задачу. Пусть на плоскости проведено n прямых, причём среди них нет параллельных, и никакие три прямые не пересекаются в одной точке. Такое положение прямых на плоскости называется их общим положением. Плоскость разбивается прямыми на некоторое число областей. Требуется доказать, что число областей равно $\frac{n^2+n+2}{2}$. На рис. 38 показано разбиение плоскости тремя прямыми общего положения на 7 областей.

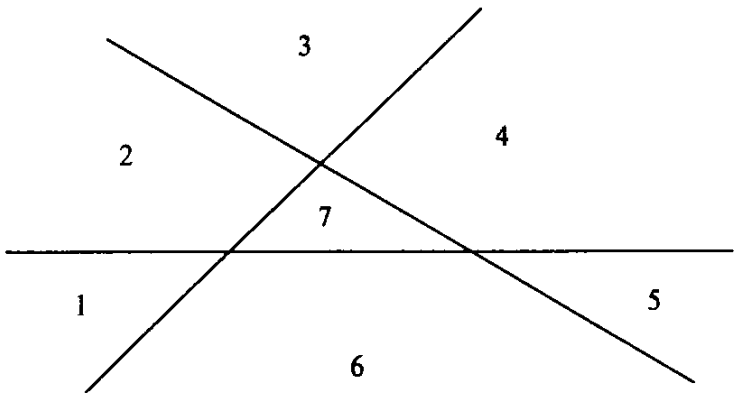


Рис. 38

Утверждение справедливо при $n = 1$, так как одна прямая разбивает плоскость на две области. Пусть утверждение верно для некоторого n, т. е. любые n прямых общего положения разбивают плоскость на $\frac{n^2+n+2}{2}$ областей. Докажем, что $n+1$ прямых общего положения всегда разбивают плоскость на $\frac{(n+1)^2+(n+1)+2}{2}$ областей. Пусть имеется $n+1$ прямых общего положения. Временно сотрём одну из прямых. Оставшиеся n прямых находятся в общем положении и, следовательно, по индуктивному предположению разбивают плоскость на $\frac{n^2+n+2}{2}$ областей. Восстановив теперь стёртую прямую, замечаем, что она разбивается остальными n

прямыми на $n+1$ частей. Эти части делят на две области каждую область, через которую проходит эта прямая. Поэтому при восстановлении прямой число областей возрастёт на $n+1$, и полное число областей будет равно

$$\left(\frac{n^2+n+2}{2}\right)+(n+1)=\frac{(n+1)^2+(n+1)+2}{2}.$$

Методом математической индукции можно не только подтверждать количественные формулы, но также и доказывать теоремы существования. Приведём два поучительных примера подобного рода. Как отмечалось в предисловии, задача о раскраске произвольной карты чстырьмя красками является одной из очень трудных задач, решение которой удалось получить сравнительно недавно. Рассмотрим, однако, простейший вариант этой задачи.

◆ Раскраска областей

Пусть на плоскости произвольным образом проведено любое конечное число прямых. Они разбивают плоскость на многоугольные области (конечные или бесконечные). Требуется доказать, что эти области могут быть раскрашены двумя цветами так, чтобы никакие две соседние области не были закрашены одним цветом.

Вот пример правильной раскраски карты, полученной проведением четырёх прямых (рис. 39).

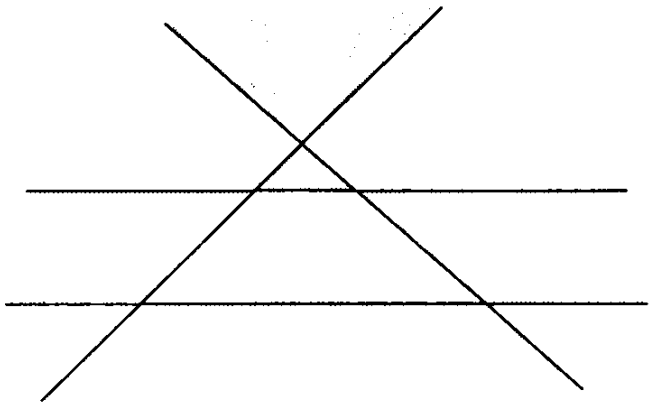


Рис. 39

Доказательство того, что правильная раскраска всегда возможна, легко проводится индукцией по числу прямых. Одна прямая разбивает плоскость на две полуплоскости, которые раскрашиваются двумя цветами, т. е. база индукции справедлива. Пусть теперь n — произвольное натуральное число. Допустив, что карта, полученная с помощью n прямых, всегда раскрашиваема в два цвета, докажем, что правильная раскраска всегда существует и для карты, полученной с помощью $n+1$ прямых.

Пусть имеется произвольная карта, полученная с помощью $n+1$ прямых. Убрав одну из прямых, получим карту с n прямыми, которая по предположению индукции может быть раскрашена двумя цветами. Возьмём такую раскраску, а теперь восстановим убранную прямую и с одной

её стороны оставим раскраску неизменной, а в другой — изменим на противоположную. Полученная таким образом раскраска будет, очевидно, правильной. Этим возможность правильной раскраски в два цвета доказана для любого числа прямых.

Автомобили на кольце ➡

Рассмотрим другой пример. Пусть на кольцевой дороге стоит произвольное число однотипных автомобилей. Известно, что суммарное количество бензина в их бензобаках достаточно для того, чтобы один автомобиль мог совершить полный круг. Нужно доказать, что для произвольно выбранного направления, по часовой или против часовой стрелки, всегда найдётся автомобиль, который, начав двигаться в данном направлении и забирая бензин по ходу движения у стоящих на дороге автомобилей, сможет совершить полный круг.

Доказательство проведём индукцией по числу автомобилей. Для одного автомобиля утверждение, очевидно, справедливо. Пусть оно справедливо для n автомобилей. Докажем его справедливость для $n+1$ автомобиля. Среди $n+1$ автомобилей всегда найдётся такой автомобиль (пусть это будет автомобиль a), который сможет доехать до ближайшего в данном направлении автомобиля (пусть это будет автомобиль b), так как в противном случае суммарного количества бензина было бы недостаточно для совершения круга одному автомобилю. Теперь мысленно уберём автомобиль b с кольцевой дороги, перелив его бензин автомобилю a. Среди оставшихся n автомобилей по предположению индукции существует автомобиль, способный проехать полный круг. Теперь, восстановив исходное положение $n+1$ автомобилей, легко видеть, что этот же автомобиль способен совершить полный круг.

Единственность разложения на множители ➡

Теперь мы в состоянии доказать по индукции единственность разложения натуральных чисел на простые множители. Пусть единственность разложения для всех чисел, не превышающих n, уже установлена. Докажем единственность разложения числа $n+1$. Если $n+1$ — число простое, то доказывать нечего. Пусть $n+1$ — составное число. Допустим, что имеется два различных разложения числа $n+1$ на простые множители:

$$n+1=p_1p_2\ldots p_k \text{ и } n+1=q_1q_2\ldots q_l,$$

где

$$p_1\le p_2\le\ldots\le p_k \text{ и } q_1\le q_2\le\ldots\le q_l.$$

Одно и то же простое число не может входить в два различных разложения, так как, сократив на него, можно было бы получить два различных разложения для числа, не превышающего n. Ясно, что $p_1q_1<n+1$. Рассмотрим число $n+1-p_1q_1$. Оно делится как на p_1, так и на q_1. А так как

его величина не превышает n, то вследствие единственности разложения оно должно делиться и на произведение $p_1 q_1$. Поэтому $n+1$ также должно делиться на $p_1 q_1$. Сокращая первое из разложений числа $n+1$ на p_1, получаем, что не превышающее n число $p_2 \ldots p_k$ имеет ещё и другое разложение на простые множители, содержащее q_1, а это противоречит индуктивному предположению.

Необходимые и достаточные условия. Если из некоторого условия A следует другое условие B, то говорят, что A является *достаточным условием* для B, и обозначают это как $A \Rightarrow B$. Условие B в этом случае называется *необходимым условием* для A, так как, если не выполнено A, то не может иметь место и B.

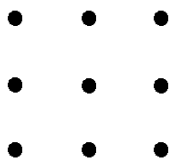
Если одновременно имеет место $A \Rightarrow B$ и $B \Rightarrow A$, то условие A называется *необходимым и достаточным условием* для B. Условие B в этом случае также является необходимым и достаточным для A. Это записывается как $A \Leftrightarrow B$. Условия A и B в этом случае могут выполняться только одновременно, поэтому их также называют *равносильными условиями*, а выражение «*необходимо и достаточно*» часто заменяют выражениями «*тогда и только тогда, когда*», «*в том и только в том случае, если*». Необходимое и достаточное условие называют также *критерием*.

Пусть, в качестве примера, A обозначает условие, состоящее в том, что некоторый треугольник является прямоугольным, а B — что в этом треугольнике сумма квадратов двух сторон равна квадрату третьей стороны. Тогда запись $A \Rightarrow B$ выражает собой теорему Пифагора «В прямоугольном треугольнике сумма квадратов катетов равна квадрату гипотенузы», а $B \Rightarrow A$ — обратную ей теорему: «Если в треугольнике сумма квадратов длин двух сторон равна квадрату длины третьей, то треугольник прямоугольный». Обратная теорема в данном случае также справедлива ввиду равенства треугольников по трём сторонам. Поэтому равенство суммы квадратов двух сторон треугольника квадрату третьей стороны необходимо и достаточно для того, чтобы треугольник был прямоугольным, или $A \Leftrightarrow B$.

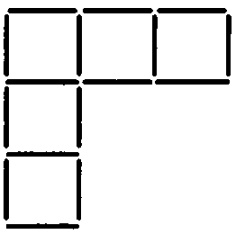
Задачи для самостоятельного решения

Все предлагаемые ниже задачи не требуют длинных вычислений и рассуждений, однако каждая из них содержит определённую «изюминку». При этом некоторые из задач являются просто тренирующими комбинаторное мышление головоломками, другие содержат в зачатке глубокие математические идеи, развиваемые далее на страницах этой книги. Решения некоторых из задач будут использованы в последующих главах книги. Спектр задач достаточно широк, чтобы каждый мог найти себе задачи по силам и по вкусу. Если решить задачу долго не удаётся, то можно посмотреть в ответ и оценить приведённое там решение.

1. В квадратной комнате расставить 5 стульев вдоль стен так, чтобы у каждой стены было одинаковое число стульев (стул, поставленный в угол, засчитывается обеим образующим угол стенам).
2. Некоторая бактерия, через секунду после появления на свет делится пополам, порождая двух таких же бактерий, которые также размножаются делением каждую секунду. Одна бактерия, посаженная в колбу, заполнила её за 1 минуту. За какое время будет заполнена колба, если в неё посадить две бактерии?
3. На неограниченно большом листе бумаги отмечено 9 точек, расположенных как указано на рисунке. Требуется перечеркнуть их четырьмя прямыми, не отрывая карандаша от бумаги и не возвращаясь назад по проведённой линии.



4. На столе лежат 16 спичек, как показано на рисунке, которые образуют 5 квадратов. Требуется переложить 2 спички так, чтобы получилось 4 таких же квадрата.



5. При заготовке дров было выполнено 20 распилов и получилось 30 поленьев. Сколько брёвен было вначале?
6. Можно ли шахматную доску с вырезанными противоположными угловыми клетками (полями a1 и h8) замостить прямоугольниками со сторонами 1×2, каждый из которых может закрывать две соседние клетки?
7. Имеются три одинаковых урны, в одной из которых 2 белых шара, в другой — белый и чёрный, в третьей — 2 чёрных. Изготовлены три соответствующие этикетки и наклеены на урны так, что ни одна из них не соответствует действительности. Как, вынув лишь один шар, узнать содержимое каждой урны?
8. Известно, что среди 9 монет имеется одна фальшивая, которая весит меньше остальных. Имеются весы, позволяющие определять, на какой из двух чаш находится больший груз. Как с помощью двух взвешиваний найти фальшивую монету?
9. Известно, что среди 9 монет имеется одна фальшивая, которая весит меньше остальных. Имеются двое весов, каждые из которых позволяют сравнивать грузы по весу. Однако одни из имеющихся весов являются недостаточно точными, чтобы улавливать разницу в весе между фальшивой и настоящей монетами. Зато другие всегда отмечают

более лёгкий груз. Но какие из весов точные, а какие нет, неизвестно. Как с помощью трёх взвешиваний найти фальшивую монету?

10. Имеется 10 мешков монет. В девяти мешках монеты настоящие (весят по 10 г.), а в одном мешке все монеты фальшивые (весят по 11 г.). Как с помощью весов, определяющих точный вес груза, с помощью одного взвешивания определить, в каком мешке фальшивые монеты?
11. Мужчины и женщины приходили на светский раут и, пробыв там некоторое время, покидали его. При этом каждый мужчина переговорил с каждой из женщин. Доказать, что в некоторый момент времени на рауте одновременно присутствовали все мужчины или все женщины.
12. Какое утверждение в следующем списке из трёх утверждений является истинным?
    1. Одно утверждение в этом списке ложно.
    2. Два утверждения в этом списке ложны.
    3. Три утверждения в этом списке ложны.
13. $m \times n$ человек построены в виде прямоугольника с m рядами и n колоннами. В каждом ряду выбирается самый высокий, а среди них берётся самый маленький. Затем в каждой колонне выбирается самый высокий, а среди них берётся самый маленький. Что можно сказать о соотношении ростов двух выбранных таким образом людей?
14. Имеется 8-литровый сосуд, до краёв наполненный водой, и два пустых объёмом 3 и 5 л. Как разлить воду поровну в два больших сосуда?
15. Полный бокал вина выплёскивается в бочку с водой, а затем снова наполняется из бочки. Чего после этого будет больше, воды в бокале или вина в бочке?
16. Вы хотите послать другу документ, содержание которого должно остаться в тайне. У вас есть коробочка с двумя парами ушек для двух навесных замков, которую вы можете посылать вашему другу, а он — вам. Подходящий замок имеется как у вас, так и у вашего друга. Но это различные замки, и ключи от одного не подходят к другому. Посылать ключ в незапертой коробочке вы не хотите, опасаясь, что он может быть скопирован. Как вы должны поступить?
17. Неизвестно, чем является начальная позиция в шахматной игре — ничьей или победой белых. Так, придворный музыкант французских королей и сильнейший шахматист XVIII века Филидор (чью защиту использовали против Остапа Бендера любители шахмат из Васюков) считал, что при правильной игре белые должны победить. Большинство же современных исследователей склонно оценивать начальную позицию как ничейную. Подобный вопрос, однако, легко решается для древней восточной игры Ним. Вот её простейший вариант.

    В куче лежит 21 камень. Двое играющих по очереди берут из кучи по желанию один, два или три камня. Выигрывает тот, кто берёт последний камень. Какой стратегии должен придерживаться делающий первый ход, чтобы победить? Исследовать задачу в общем случае,

когда в куче лежит n камней, а каждый из играющих может брать от одного до k камней.

18. Четыре человека, у которых один фонарик, должны ночью перейти мост. Одновременно по мосту могут идти не более двух человек и у них обязательно должен быть фонарик. Каждый из них переходит мост за различное время: первый — за 1 минуту, второй — за 2, третий — за 5 и четвёртый — за 10 минут. Если идут вдвоём, то время их движения определяется временем более медленного из них. Фонарик нельзя перебрасывать с одного берега на другой, а можно лишь переносить обратно. Как должны действовать эти четверо, чтобы им всем перейти мост за 17 минут? (Эту задачу можно отнести к математической теории расписаний!)
19. Доказать, что $1\cdot 1!+2\cdot 2!+\ldots+n\cdot n!=(n+1)!-1$.
20. Доказать, что $9^{n+1}-8n-9$ делится на 64 при любом натуральном n.
21. Как найти a^{10}, сделав лишь четыре умножения? (промежуточные результаты можно запоминать)
22. Доказать, что для любых натуральных m и n их наибольший общий делитель (m, n) и наименьшее общее кратное $[m, n]$ связаны соотношением $(m,n)\cdot[m, n]=m\,n$.
23. Некто купил лошадей и быков на общую сумму в 1770 талеров. За каждую лошадь он уплатил по 31 талеру, а за каждого быка — по 21 талеру. Что можно сказать о числе купленных им лошадей и быков?
24. Пусть $n=p\,q$, где p и q два различных простых числа. Доказать, что $\varphi(pq)=(p-1)(q-1)$.
25. Пусть снова $n=p\,q$, где p и q два различных простых числа. Доказать, что из $a\equiv b(\mathrm{mod}\,p)$ и $a\equiv b(\mathrm{mod}\,q)$ следует, что $a\equiv b(\mathrm{mod}\,n)$.

Этот результат будет использован при доказательстве корректности криптографической системы RSA в разделе 5.12. Он является частным случаем так называемой Китайской теоремы об остатках, которая формулируется следующим образом:

если $m_1, m_2, \ldots, m_n$ *— попарно взаимно простые числа и для чисел* a *и* b *выполнены сравнения*

$$a\equiv b\ (\mathrm{mod}\,m_1),$$
$$a\equiv b\ (\mathrm{mod}\,m_2),$$
$$\vdots$$
$$a\equiv b\ (\mathrm{mod}\,m_n),$$

то справедливо сравнение

$$a\equiv b\ (\mathrm{mod}\,m),$$

где $m=m_1m_2\ldots m_n$.

26. Доказать, что если 2^n+1 простое число, то n — степень двойки. (Указание: использовать справедливое для нечётного k тождество

$$a^k+b^k=(a+b)(a^{k-1}-a^{k-2}b+a^{k-3}b^2-\ldots-ab^{k-2}+b^{k-1}).)$$

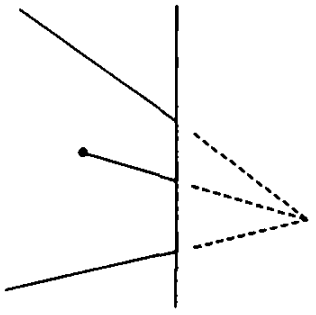
Числа вида $2^{2^k}+1$ называют числами Ферма. Пьер Ферма предполагал, что для любого $k=0,1,2,\ldots$ числа такого вида являются простыми. Это предположение, подтверждалось тем, что первые 5 чисел этого вида $3, 5, 17, 257, 65\,537$ действительно оказываются простыми.

Но, как было показано спустя столетие Эйлером, $2^{2^5}+1$ является составным числом:

$$2^{2^5}+1=4\,294\,967\,297=641\times 6\,700\,417.$$

Более того, ни одного простого числа Ферма, за исключением перечисленных пяти, найти, пока, не удалось.

27. Доказать, что любые два числа в последовательности чисел Ферма $2^{2^k}+1$, $k=0,1,2,\ldots$ взаимно просты.
28. Пусть $A_1A_2\ldots A_n$ — правильный n-угольник, точка O — его центр. Доказать, что $\overrightarrow{OA_1}+\overrightarrow{OA_2}+\ldots+\overrightarrow{OA_n}=\vec{0}$.
29. Доказать, что если ненулевые векторы $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ попарно ортогональны, то они линейно независимы.
30. Пусть S и T — подпространства векторного пространства. Доказать, что $(S+T)^{\perp}=S^{\perp}\cap T^{\perp}$.
31. На ограниченном с одной стороны чертеже имеется точка и две прямые, пересекающиеся за пределами чертежа. Как, используя симметрию, через эту точку провести прямую, продолжение которой за пределы чертежа пройдёт через точку пересечения двух прямых?



32. Показать, что любая подгруппа циклической группы также является циклической группой.
33. Показать, что транспозиция (12) и цикл $(123\ldots n)$ порождают всю симметрическую группу S_n.
34. Показать, что конечное подмножество элементов группы, замкнутое относительно операции группового умножения, является подгруппой,

т. е. единичный элемент и все обратные элементы также принадлежат этому подмножеству.

35. Доказать, что если в некоторой группе квадрат любого элемента есть единичный элемент, то это коммутативная группа.
36. Элементы a и b коммутативной группы, имеют соответственно порядки m и n, которые взаимно просты. Доказать, что порядок их произведения ab равен mn.
37. В экзаменационную сессию поток студентов сдавал 4 предмета: математику, физику, химию и иностранный язык. Математику сдали на «отлично» 70 % студентов, физику — 75 %, химию — 80 % и иностранный язык — 85 %. В каком интервале может лежать процент студентов, получивших повышенную стипендию (сдавших все экзамены на «отлично»)?
38. Сколько диагоналей можно провести в правильном n-угольнике? (Диагональю называется отрезок, соединяющий любые две несмежные вершины.)
39. Назовём многозначное число симметричным, если оно не изменяет своего значения при чтении в обратном порядке, например, 33, 252, 6446 и т. д. Сколько существует симметричных 7-значных чисел?
40. Требуется составить расписание для проведения 5 уроков: по алгебре, геометрии, физике, химии и биологии с тем условием, чтобы алгебра и геометрия не шли подряд. Сколькими способами это можно сделать?
41. За круглым столом нужно рассадить n мужчин и n женщин так, чтобы никакие два лица одного пола не сидели рядом. Сколькими способами это можно сделать, если считать расположения, получаемые круговым сдвигом всей компании, одинаковыми?
42. Из букв разрезной азбуки составляется слово «катамаран». Затем 9 использованных в слове карточек перемешиваются и снова собираются в случайном порядке. Какова вероятность, что снова возникнет слово «катамаран»?
43. Документ лежит в столе с вероятностью $1/2$, а с вероятностью $1/2$ находится в другом месте. Находясь в столе, он может с равной вероятностью лежать в одном из его четырёх ящиков. Три ящика стола были просмотрены, и в них документа не оказалось. Какова вероятность обнаружить его в четвёртом ящике?
44. Осьминог Пауль, использовавшийся в качестве оракула во время чемпионата 2010 года по футболу, правильно предсказал исходы всех 8 матчей, которые с его помощью пытались угадать. Если считать выбор осьминога между двумя кормушками, помеченными флагами играющих сборных, случайным, подобно результату бросания монетки, то какова вероятность такой серии верных предсказаний?
45. Петя и Вова играют в игру, заключающуюся в следующем. Монетка подбрасывается до тех пор, пока не наберётся 10 выпадений «орла»

или 10 выпадений «решки». Если первым произойдёт 10 выпадений «орла», то выигрывает Петя, а, если «решки», — то Вова. Игра была прервана, когда «орёл» выпал 8 раз, а «решка» — 9. Какова вероятность победы для каждого из участников при продолжении игры?

46. На окружности случайным образом независимо выбираются три точки. Какова вероятность, что центр окружности окажется внутри образованного ими треугольника?

    Попробовать решить задачу, если на окружности выбираются n точек, и требуется найти вероятность того, что центр окружности окажется внутри образованного ими выпуклого многоугольника.

47. Двое договорились о встрече в условленном месте в 6 часов. Каждый из них, не будучи слишком точным, может прийти в любое время с 6 до 7. Придя на место встречи, любой ждёт другого в течение получаса, а не дождавшись в течение этого времени, уходит (разумеется, до встречи «завтра, на том же месте и в тот же час»). Какова же вероятность встречи сегодня?

48. Для оценки количества рыб, обитающих в озере, было выловлено, помечено и снова выпущено в озеро n_1 рыб. Затем n_2 рыб последовательно вылавливались и отпускались, и среди них оказалось m помеченных. Воспользовавшись вероятностной моделью, оценить число рыб в озере.

49. Два корректора независимо вычитывают один и тот же текст. Один находит n_1 ошибок, другой — n_1, причём n_{12} ошибок было замечено обоими корректорами. Воспользовавшись вероятностной моделью, оценить число ошибок, оставшихся незамеченными.

50. Король вызвал капитана королевских мушкетёров Д'Артаньяна и спросил его, есть ли среди его мушкетёров ссоры. Д'Артаньян честно ответил, что каждый из его мушкетёров находится в ссоре с тремя другими мушкетёрами. Тогда для предотвращения дуэлей король приказал д'Артаньяну разделить мушкетёров на несколько отрядов так, чтобы внутри каждого отряда не было поссорившихся. Какое минимальное число отрядов гарантирует достижение этой цели?

51. Задан произвольный граф. Нужно расставить в его вершинах натуральные числа так, чтобы для каждой пары вершин эти числа были бы взаимно просты, если вершины несмежны, и имели бы отличный от единицы общий делитель, если вершины смежны. Как этого всегда можно легко добиться?

52. Доказать, что в любой компании всегда найдутся двое, имеющие одинаковое число знакомых среди лиц данной компании.

53. Доказать, что в произвольной группе из 6 человек всегда найдутся трое попарно знакомых между собой или трое попарно незнакомых между собой.

54. Доказать, что любую группу людей всегда можно развести по двум комнатам так, что у каждого в другой комнате было не менее половины из его знакомых в данной компании.

55. Плоскости в пространстве находятся в общем положении, если каждая пара плоскостей пересекается по прямой, каждая тройка плоскостей имеет одну общую точку, а пересечение каждой четвёрки плоскостей пусто. Показать, что n плоскостей общего положения разбивают пространство на $\frac{n^3+5n+6}{6}$ областей.

Литература

Приводимая ниже литература в духе издававшейся в СССР на протяжении многих лет библиотечки школьного математического кружка может служить элементарным введением в ряд разделов дискретной математики. Из неё заинтересованный читатель может извлечь для себя дальнейшие сведения по затронутым во Вводной главе темам, ограничиваясь при этом элементарным уровнем изложения. Ныне, когда в школьную программу возвращается комбинаторика и вводятся элементы теории вероятностей, этот список литературы может оказаться полезным как для школьников, так и для преподавателей.

1. *Александров П. С.* Введение в теорию групп. 3-е изд. М.:URSS, 2010.
2. *Алексеев В. Б.* Теорема Абеля в примерах и задачах. М.: Наука, 1976.
3. *Аршинов М. Н., Садовский Л. Е.* Коды и математика. М.: Наука, 1983.
4. *Беран А.* Упорядоченные множества. М.: Наука, сер. «Популярные лекции по математике». Вып. 55.
5. *Вейль Г.* Симмстрия. 3-е изд. М.: Издательство ЛКИ/URSS, 2007.
6. *Виленкин Н. Я., Виленкин А. Н., Виленкин П. А.* Комбинаторика. М: «ФИМА»-МЦНМО, 2006.
7. *Воробьёв Н. Н.* Числа Фибоначчи. М.: Наука, сер. «Популярные лекции по математике». Вып. 6.
8. *Дориченко С. А., Ященко В. В.* 25 этюдов о шифрах. М.: ТЕИС, 1994.
9. *Дынкин Е. Б., Успенский В. А.* Математические беседы. М.: Физматлит, 2004.
10. *Дэвенпорт Г.* Высшая арифметика. М.: Наука, 1965.
11. *Ежов И. И., Скороход А. В., Ядренко М. И.* Элементы комбинаторики. М.: Наука, 1977.
12. *Кемени Дж., Снелл Дж., Томпсон Дж.* Введение в конечную математику. М.: Мир, 1965.
13. *Маркушевич А. И.* Возвратные последовательности. М.: Наука, сер. «Популярные лекции по математике». Вып. 1.
14. *Оре О.* Графы и их применение. 4-е изд. М:.Издательство ЛКИ/URSS, 2008.
15. *Пойа Д.* Как решать задачу. 4-е изд. М: Книжный дом «Либроком»/URSS, 2010.
16. *Пойа Д.* Математика и правдоподобные рассуждения. 3-е изд. М: Книжный дом «Либроком»/URSS, 2010.
17. *Пойа Д.* Математическое открытие. 3-е изд. М.: КомКнига/URSS, 2010.
18. *Радемахер Г., Теплиц О.* Числа и фигуры. 4 изд. М.: Наука, 1966.
19. *Соминский И. С., Головина Л. И., Яглом И. М.* О математической индукции. М.: Наука, 1967.
20. *Фрид Э.* Элементарное введение в абстрактную алгебру. М.: Мир, 1979.
21. *Шафаревич И. Р.* Избранные главы алгебры. М.: Журнал «Математическое образование», 2000.
22. *Шашкин Ю. А.* Эйлерова характеристика. М.: Наука, сер. «Популярные лекции по математике». Вып. 58.

Глава 1

Методы перечисления[1]

1.1. Комбинаторные числа

Перечислительная комбинаторика отвечает на вопрос «сколько?» и занимается подсчётом числа объектов, построенных по определённым правилам из заданного конечного множества элементов. Простейшими такими объектами являются перестановки, размещения, сочетания и разбиения. Заметим, что элементарная комбинаторика почти не использует специальных математических знаний, но требует концентрации внимания и определённой остроты ума, как при игре в шахматы. Но для того, чтобы играть в шахматы, нужно освоить правила и познакомиться со стандартными комбинациями. Поэтому начнём с правил и сделаем первые ходы.

Перестановки. Пусть задано конечное множество из n элементов, которые будем называть символами. В качестве таких символов могут выступать, например, числа $\{1, 2, \dots, n\}$. *Перестановкой* на n символах называется последовательность n символов, выписанных в определённом порядке. Таким образом, число перестановок на n символах, обозначаемое P_n — это число способов введения линейного порядка на множестве из n элементов. Можно сказать также, что P_n — это число способов расставить n человек в очередь. На первое место можно поставить любого из n, на второе — любого из оставшихся $n-1$ и т. д. пока не дойдем до n-го места, на которое останется единственный представитель. Поэтому

$$P_n = n\cdot(n-1)\cdot\ldots\cdot 2\cdot 1 = n!.$$

Напомним, что

$$0! = 1! = 1,$$

$$2! = 1\cdot 2 = 2,\quad 3! = 1\cdot 2\cdot 3 = 2!\cdot 3 = 6.$$

---

[1] Нумерация утверждений, лемм, теорем, рисунков и таблиц начинается заново в каждом разделе каждой главы. При ссылке внутри раздела указывается лишь внутренний номер, при ссылке внутри главы—номер раздела и внутренний номер, а при ссылке из другой главы—номер главы, раздела и внутренний номер.

Заметим также, что $n! = n \cdot (n-1)!$. Формулы подобного типа, выражающие функцию от натурального аргумента через её значения при меньших значениях аргумента, называются *рекуррентными*.

Размещения. Пусть теперь из n элементов требуется выбрать m элементов $(m \le n)$ и линейно их упорядочить. Обозначив число таких упорядоченных выборок через A_n^m и рассуждая как и прежде, получаем

$$A_n^m = n(n-1)\ldots(n-m+1) =$$

$$= \frac{n(n-1)\ldots(n-m+1)(n-m)(n-m-1)\cdot\ldots\cdot 2\cdot 1}{(n-m)(n-m-1)\cdot\ldots\cdot 2\cdot 1} = .$$

$$= \frac{n!}{(n-m)!}$$

Число A_n^m называют также *числом размещений из n по m*, так как m элементов из n размещаются по m местам, пронумерованным от 1 до m[1]. Полагают $A_n^0 = 1$ и $A_n^m = 0$ при $m > n$. Вот множество упорядоченных выборок длины 2 из 3-элементного множества: $\{(1,2), (2,1), (1,3), (3,1), (2,3), (3,2)\}$, мощность которого в соответствии с выведенной формулой равна $A_3^2 = \frac{3!}{1!} = 6$.

Сочетания. Пусть, наконец, из n-элементного множества просто выбирается его m-элементное подмножество без упорядочивания. Число m-элементных подмножеств n-элементного множества является весьма важным не только для комбинаторики, но и для всей математики. Оно обозначается через C_n^m (читается, «це» из n по m) и называется в комбинаторике *числом сочетаний из n по m*. В других разделах математики по причинам, которые прояснятся чуть позже, их обычно называют биномиальными коэффициентами[2].

Для нахождения C_n^m заметим, что упорядоченную выборку можно рассматривать как получаемую в два этапа: сначала из n элементов выбирает-

---

[1] В научной литературе вместо A_n^m может использоваться обозначение $(n)_m$.

[2] В научной литературе вместо C_n^m может использоваться обозначение $\binom{n}{m}$.

ся неупорядоченное m-элементное подмножество, что можно сделать C_n^m способами, а затем выбранное m-элементное подмножество линейно упорядочивается, что можно сделать $P_m = m!$ способами. Это приводит к соотношению:

$$A_n^m = C_n^m \cdot m!,$$

откуда получаем

$$C_n^m = \frac{A_n^m}{m!} = \frac{n!}{(n-m)! \cdot m!}.$$

Пусть имеется 4-элементное множество $\{a,b,c,d\}$. Вот все его двухэлементные подмножества: $\{a,b\},\{a,c\},\{a,d\},\{b,c\},\{b,d\},\{c,d\}$, число которых в соответствии с формулой равно $C_4^2 = \frac{4!}{2! \cdot 2!} = 6$.

Запомним, что $C_n^m = 0$, если $m > n$ или $m < 0$.

Из формулы для числа сочетаний непосредственно вытекают также следующие равенства

$$C_n^m = C_n^{n-m} \text{ и } C_n^0 = C_n^n = 1,\ C_n^1 = C_n^{n-1} = n,\ C_n^2 = C_n^{n-2} = \frac{n(n-1)}{2}.$$

Проиллюстрируем введенные комбинаторные числа простейшими примерами.

> Пусть в спортивном тотализаторе требуется указать золотого, серебряного и бронзового призёров из 16 команд высшей лиги. Сколькими способами это может быть сделано?

Число способов есть $A_{16}^3 = \frac{16!}{13!} = \frac{13! \cdot 14 \cdot 15 \cdot 16}{13!} = 14 \cdot 15 \cdot 16 = 3360$. ◀

> Сколькими способами студенческая группа в составе 20 человек может делегатировать трёх человек для участия в студенческой конференции?

Число способов равно $C_{20}^3 = \frac{20!}{17! \cdot 3!} = \frac{18 \cdot 19 \cdot 20}{1 \cdot 2 \cdot 3} = 1140$. ◀

> Если футбольный матч закончился вничью и его судьба решается в серии послематчевых пенальти, то сколькими способами тренер может представить судье список 5 пенальтистов из 11 футболистов, находившихся на поле в момент финального свистка?

Число способов есть $A_{11}^5 = \frac{11!}{6!} = 7 \cdot 8 \cdot 9 \cdot 10 \cdot 11 = 55440$, так как порядок выполнения футболистами пенальти имеет значение. ◄

Сколькими способами может быть перетасована 36-карточная колода?

$$P_{36} = 36!.\blacktriangleleft$$

Сколькими способами можно рассадить 200 человек в зрительном зале, имеющем 250 нумерованных мест?

$$A_{250}^{200} = \frac{250!}{50!}.\blacktriangleleft$$

Правило произведения

Пусть некоторый выбор осуществляется в два этапа, причём на первом этапе имеется k_1 возможностей для выбора, на втором — k_2. Тогда полное число возможных вариантов выбора равно $k_1 k_2$ ➡

При нахождении чисел P_n, A_n^m, C_n^m фактически постоянно использовалось «правило произведения», которое теперь сформулировано в явном виде.

Приведём дополнительные наглядные примеры применения этого правила.

Из города A в город B ведут 2 дороги, а из города B в город C — 3 дороги. Сколько существует маршрутов, ведущих из города A в город C через город B?

Существует $2 \cdot 3 = 6$ маршрутов. ◄

Ресторанное меню содержит 7 первых и 8 вторых блюд. Сколькими способами может быть сделан заказ из первого и второго блюда?

Число возможных заказов равно $7 \cdot 8 = 56$.◄

Тренер футбольной команды желает одновременно заменить троих из 10 полевых игроков, имея 5 футболистов на скамейке запасных. Сколькими способами он может это сделать?

$C_{10}^3 \cdot C_5^3 = \frac{10 \cdot 9 \cdot 8}{2 \cdot 3} \cdot \frac{5 \cdot 4 \cdot 3}{2 \cdot 3} = 1200$ способами, на первом этапе выбирая трёх покидающих поле футболистов, а на втором — трёх футболистов, выходящих на замену. ◄

Правило произведения естественным образом обобщается на случай, когда выбор производится не в два, а в большее число этапов. Полное число

вариантов выбора равно произведению количеств возможностей на всех этапах. Таким образом, если выбор производится в n этапов и на i-м этапе имеется k_i возможностей, то полное число возможностей равно $k_1 k_2 \ldots k_n$.

Некто, имея достаточно денег, зашёл в буфет, где имеется 5 ванильных, 6 шоколадных и 7 фруктовых пирожных. Сколько вариантов сладкой покупки у него есть, включая и вариант ничего не купить?

$(5 + 1)(6 + 1)(7 + 1) = 336$ вариантов. Здесь на трёх этапах выбора определяются количества покупаемых пирожных каждого типа. ◀

Электропоезд с 10 пустыми пассажирскими вагонами подходит к платформе, на которой его ожидают 20 человек. Сколькими способами они могут разместиться по вагонам?

10^{20} способами, так как для каждого человека существует 10 возможностей для выбора вагона. ◀

Не будет преувеличением сказать, что правило произведения является основой элементарной комбинаторики. Его можно наглядно представить с помощью дерева выбора возможностей (рис. 1).

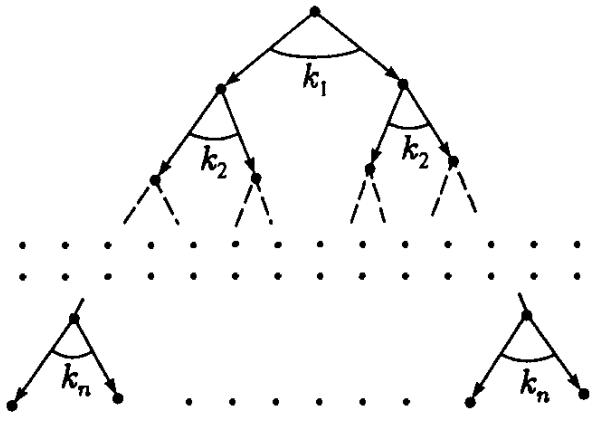


Рис. 1

При его применении нужно лишь следить за тем, чтобы при ветвлении в каждой вершине происходило разбиение множества вариантов на непересекающиеся подмножества. Этим гарантируется, что окончательные варианты, получаемые на концах ветвей дерева, будут различными при движении по различным ветвям. Поясним это на примере. Пусть с помощью правила произведения хотят найти число двухэлементных подмножеств n-элементного множества, считая, что на первом этапе формирования подмножества выбирается один элемент, а на втором — другой. Тогда на первом этапе выбор производится из n возможностей, а на втором — из $n-1$. Поэтому в соответствии с правилом произведения получилось бы, что число двухэлементных подмножеств равно $n(n-1)$, что неверно, так как это число равно $C_n^2 = n(n-1)/2$. Ошибка объясняется тем, что при таком порождении двухэлементных подмножеств каждое подмножество возникает дважды: при одном порядке включения входящих в него элементов и при другом. И здесь с помощью

правила произведения на самом деле находится число упорядоченных двухэлементных выборок A_n^2.

Последовательности символов. Рассмотрим множество двоичных наборов длины n, т. е. множество последовательностей $(a_1, a_2, \ldots, a_n)$ [1], где $a_i \in \{0,1\}$. Так как на каждое место набора можно независимо ставить 0 или 1, то всего наборов 2^n. Такие наборы называют также бинарными. Считая их характеристическими векторами подмножеств n-элементного множества, заключаем, что полное число подмножеств n-элементного множества также равно 2^n.

Число различных наборов $(a_1, a_2, \ldots, a_n)$ длины n, компоненты a_i которых выбираются из m-элементного множества символов $\{c_1, c_2, \ldots, c_m\}$ равно m^n, так как на каждое место можно независимо ставить любой из m символов. В частности, число целых неотрицательных не более чем n-значных десятичных чисел равно 10^n.

Рассмотренные множества двоичных и m-значных наборов длины n являются декартовыми степенями соответствующих множеств — $\{0,1\}^n$ и $\{c_1, c_2, \ldots, c_m\}^n$. Их мощность можно было бы также получить из формулы для мощности прямого произведения

$$|A \times B| = |A| \cdot |B|,$$

которая также является прямым следствием комбинаторного правила произведения.

Рассмотрим теперь множество наборов $(a_1, a_2, \ldots, a_n)$, где $a_i \in \{c_1, c_2, \ldots, c_m\}$, таких, что в каждом наборе символ c_1 встречается n_1 раз, символ c_2 — n_2 раз и т. д., символ c_m — n_m раз $(n_1 + n_2 + \ldots + n_m = n)$. Чему равно число таких наборов из m символов с заданной кратностью n_i использования каждого символа c_i в наборе? Руководствуясь правилом произведения, можно поступить следующим образом. На первом этапе из n возможных мест выбираем n_1 мест для символа c_1, для чего есть $C_n^{n_1}$ возможностей. На втором этапе из $n - n_1$ оставшихся свободными мест выбираем n_2 мест для символа c_2, что осуществимо $C_{n-n_1}^{n_2}$ способами. На третьем из $n - n_1 - n_2$ мест выбираем n_3 мест для символа c_3 и т. д. В итоге получаем, что искомое число последовательностей равно

---

[1] Подобные наборы называют также кортежами. Здесь это несколько напыщенное французское слово использоваться не будет.

$$C_n^{n_1} C_{n-n_1}^{n_2} C_{n-n_1-n_2}^{n_3} \dots C_{n_m}^{n_m} =$$

$$= \frac{n!}{n_1!(n-n_1)!} \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \frac{(n-n_1-n_2)!}{n_3!(n-n_1-n_2-n_3)!} \dots \frac{n_m!}{n_m!} = \frac{n!}{n_1!n_2!\cdots n_m!}.$$

К решению этой задачи возможен, однако, и другой, более изящный подход. Возьмём n карточек и напишем на них n_1 символов c_1, n_2 символов c_2, и т. д., n_m символов c_m. Тогда каждой перестановке n карточек будет соответствовать набор, включающая n_1 раз символ c_1, n_2 раз символ c_2 и т. д. Однако перестановки карточек с одинаковыми символами не изменяют набора. Таких перестановок $n_1!\cdot n_2!\cdot \dots \cdot n_m!$. Отсюда получаем:

> Количество наборов длины n из m символов с заданной кратностью n_i использования каждого символа равно
>
> $$\frac{n!}{n_1!\cdot n_2!\cdot \dots \cdot n_m!}.$$

В частности, число двоичных последовательностей длины n с m единицами равно $\frac{n!}{m!(n-m)!} = C_n^m$. Оно совпадает с числом m-элементных подмножеств n-элементного множества, так как каждый такой набор можно рассматривать как характеристический вектор m-элементного подмножества.

В качестве примера рассмотрим наборы (a_1, a_2, a_3, a_4), где $a_i \in \{\alpha, \beta, \gamma\}$. Полное число таких наборов равно $3^4 = 81$. Найдём число наборов, в которых символ α встречается дважды, а символы β и γ по одному разу. В соответствии с формулой число таких наборов равно $\frac{4!}{2!\cdot 1!\cdot 1!} = 12$. Вот эти 12 наборов: $(\alpha\alpha\beta\gamma)$, $(\alpha\alpha\gamma\beta)$, $(\alpha\beta\alpha\gamma)$, $(\alpha\gamma\alpha\beta)$, $(\alpha\beta\gamma\alpha)$, $(\alpha\gamma\beta\alpha)$, $(\beta\alpha\alpha\gamma)$, $(\gamma\alpha\alpha\beta)$, $(\beta\alpha\gamma\alpha)$, $(\gamma\alpha\beta\alpha)$, $(\beta\gamma\alpha\alpha)$, $(\gamma\beta\alpha\alpha)$.

Сочетания с повторениями. Пусть теперь имеется n типов элементов, где элементы одного типа считаются неразличимыми, а запас элементов каждого типа неограничен. Рассматриваются m-элементные множества, в каждое из которых может включаться любое количество элементов каждого типа от 0 до m. Число таких множеств называется *числом сочетаний с повторениями из n по m* и обозначается $\overline{C}_n^m$. Здесь возможен случай и $m > n$.

Сочетание с повторениями называют также *мультимножеством*. Если обычное множество задаётся характеристическим вектором, компоненты

которого принимают значения 0 или 1, то мультимножество на n типах элементов может быть задано *вектором кратностей типов* $(\lambda_1, \lambda_2, \ldots, \lambda_n)$, где λ_i — число элементов i-го типа в мультимножестве, которое может быть любым целым неотрицательным числом. Мощность мультимножества A определяется как сумма кратностей типов: $|A| = \sum_{i=1}^{n} \lambda_i$. В терминах мультимножества $\overline{C}_n^m$ есть число мультимножеств мощности m, построенных на n типах элементов.

В качестве примера снова рассмотрим задачу с пирожными, слегка видоизменив её.

> В кондитерской имеются в неограниченном количестве 3 вида пирожных — ванильные, шоколадные и фруктовые. Некто желает купить 8 пирожных. Сколькими способами он может это сделать?

В наших обозначениях это число равно $\overline{C}_3^8$. ◄

> Другой пример. Желая преподнести любимой девушке букет из 5 прекрасных свежих роз, молодой человек заходит в цветочный магазин, где есть розы (прямо из Голландии!) 4 сортов: красные, жёлтые, оранжевые и белые. Сколькими способами можно выбрать букет?

Число способов есть $\overline{C}_4^5$. ◄

> А вот пример из военной области. Представьте, что Вы — командир 6-пушечной артиллерийской батареи, на которой осталось 4 снаряда. Сколькими способами можно произвести 4 последних выстрела?

Распределить 4 снаряда по 6 орудиям можно $\overline{C}_6^4$ способами. ◄

И, наконец, в качестве примера того, как мультимножества возникают в математике, рассмотрим задачу о **числе целочисленных неотрицательных решений уравнения**

$$x_1 + x_2 + \ldots + x_n = m, \qquad x_i \geq 0, \quad i = 1, 2, \ldots, n.$$

Связывая с каждой переменной некоторый тип и считая её значение — числом элементов данного типа, т. е. рассматривая набор $(x_1, x_2, \ldots, x_n)$ как вектор кратностей типов, получаем, что число искомых решений есть число мультимножеств мощности m, построенных на n типах, т. е. равно $\overline{C}_n^m$.

Сделав эти предварительные замечания, поясняющие смысл понятия сочетаний с повторениями, найдём теперь числа $\overline{C}_n^m$. С этой целью рассмотрим наборы длины $(n+m-1)$, компонентами которых являются символы «*» и «|», причём число звездочек равно m, а число чёрточек — $(n-1)$. С каждым таким набором можно связать сочетание с повторениями, ставя в каждом из n промежутков между чёрточками вместо звездочек символы типа, соответствующего номеру промежутка. Этим устанавливается взаимно однозначное соответствие между m-элементными множествами из n типов элементов и двухсимвольными наборами длины $(n+m-1)$. Так в примере с розами покупке двух красных, одной желтой, ни одной оранжевой и двух белых роз соответствует набор $**|*||**$.

Установленное взаимно однозначное соответствие позволяет найти числа $\overline{C}_n^m$:

$$\overline{C}_n^m = C_{n+m-1}^m = \frac{(n+m-1)!}{m!(n-1)!}.$$

В качестве примера применения этой формулы рассмотрим задачу:

Найти число костяшек домино.

Кость домино состоит, как известно, из двух полей, каждое из которых помечено одним из 7 чисел от 0 до 6, причём числа на обоих полях могут совпадать (дубль). Поэтому число костей домино равно $\overline{C}_7^2 = C_8^2 = \frac{8!}{6!\cdot 2!} = 28$. ◀

Разбиения. Пусть теперь n-элементное множество разбивается на классы среди которых j_1 одноэлементных, j_2 двухэлементных, j_3 трёхэлементных и т. д., где $1j_1 + 2j_2 + \ldots + nj_n = n$. Назовём такое разбиение *разбиением типа* $(j_1, j_2, \ldots, j_n)$ и найдём число разбиений заданного типа.

С каждой перестановкой исходного n-элементного множества можно связать его разбиение типа $(j_1, j_2, \ldots, j_n)$, если в перестановке отсчитывать слева направо j_1 раз по одному элементу, j_2 раз по два элемента и т. д. При этом перестановка элементов внутри каждого класса, а также перестановка целиком классов одинаковой мощности между собой не меняют разбиения. Поэтому **число разбиений типа** $(j_1, j_2, \ldots, j_n)$ **равно**

$$\frac{n!}{\prod_{i=1}^{n} (i!)^{j_i} j_i!}. \qquad (1)$$

Число разбиений 4-элементного множества на два 2-элементных (разбиение типа (0, 2, 0, 0)) в соответствии с этой формулой равно $\frac{4!}{(2!)^2 2!} = 3$.

Вот все возможные разбиения множества $\{a,b,c,d\}$ на два двухэлементных класса: $\{a,b\} \cup \{c,d\}$, $\{a,c\} \cup \{b,d\}$, $\{a,d\} \cup \{b,c\}$.

Подчеркнём, что разбиение — это множество, элементами которого являются классы, и два разбиения одинаковы, если между их классами можно установить взаимно однозначное соответствие, при котором друг другу соответствуют одинаковые классы.

Наряду с обычными разбиениями часто рассматриваются и так называемые *упорядоченные разбиения* типа $(j_1, j_2, \ldots, j_n)$, в которых классы линейно упорядочены в соответствии с неубыванием их мощностей. В упорядоченном разбиении перестановка классов одинаковой мощности приводит к различным упорядоченным разбиениям, и **число упорядоченных разбиений типа $(j_1, j_2, \ldots, j_n)$ оказывается равным**

$$\frac{n!}{\prod_{i=1}^{n} (i!)^{j_i}}. \qquad (2)$$

Рассмотрим примеры унорядоченных и неупорядоченных разбиений.

> Сколькими способами колоду 36 карт можно разбить на две равные части?

Число разбиений равно $\frac{36!}{(18!)^2 \cdot 2!}$ (неупорядоченные разбиения). ◄

> Сколькими способами колоду 36 карт можно разделить поровну между двумя играющими?

Число разбиений равно $\frac{36!}{(18!)^2}$ (упорядоченные разбиения). ◄

> Сколькими способами из 18 человек можно составить 3 волейбольные команды по 6 человек в каждой?

Здесь количество способов есть число неупорядоченных разбиений, которое, согласно (1), равно $\frac{18!}{(6!)^3 3!}$. ◄

> Сколькими способами 18 студентов можно распределить между 3 преподавателями так, чтобы у каждого преподавателя было по 6 студентов?

Искомое число есть число упорядоченных разбиений 18-элементного множества на три 6-элементных класса, так как перестановки 6-элементных классов между собой меняют преподавателей у подгрупп. Поэтому, в соответствии с (2), это число равно $\frac{18!}{(6!)^3}$. ◀

Числа Стирлинга и Белла. *Числом Стирлинга второго рода* S_n^m называется число разбиений n-элементного множества на m непустых классов. При этом полагается $S_n^m = 0$, если $m > n$, и $S_n^0 = 0$ при $n \geq 1$, но $S_0^0 = 1$. Согласно (1) S_n^m может быть найдено как

$$S_n^m = \sum_{\substack{j_1,\ldots,j_n \geq 0 \\ 1j_1+\ldots+nj_n=n \\ j_1+\ldots+j_n=m}} \frac{n!}{\prod_{i=1}^{n} (i!)^{j_i}\, j_i!}, \tag{3}$$

Это слишком сложная формула. Гораздо удобнее вычислять S_n^m с помощью следующего рекуррентного соотношения

$$S_n^m = mS_{n-1}^m + S_{n-1}^{m-1}. \tag{4}$$

В справедливости (4) можно убедиться следующим образом. Разбиение n-элементного множества на m классов можно получить, убрав один из элементов из множества и действуя затем одним из двух способов: разбив оставшееся $(n-1)$-элементное множество на m классов и добавив убранный элемент к одному из классов, что может быть сделано mS_{n-1}^m способами, либо образовав из убранного элемента отдельное подмножество, разбив $(n-1)$-элементное множество на $m-1$ классов.

Значения S_n^m могут быть также найдены по *формуле Стирлинга*

$$S_n^m = \frac{1}{m!}\sum_{i=1}^{m}(-1)^{m-i}C_m^i\, i^n, \tag{5}$$

доказательство которой приведено в разделе 3.

Полное число всевозможных разбиений n-элементного множества называется *числом Белла* и обозначается $B(n)$. По определению имеем

$$B(n) = \sum_{m=1}^{n} S_n^m.$$

Вот все 15 разбиений четырёхэлементного множества $\{a,b,c,d\}$:

$\{a,b,c,d\}$; $\{a\}\cup\{b,c,d\}$; $\{a\}\cup\{b\}\cup\{c,d\}$; $\{a\}\cup\{b\}\cup\{c\}\cup\{d\}$.
$\{b\}\cup\{a,c,d\}$; $\{a\}\cup\{c\}\cup\{b,d\}$;
$\{c\}\cup\{a,b,d\}$; $\{a\}\cup\{d\}\cup\{b,c\}$;
$\{d\}\cup\{a,b,c\}$; $\{b\}\cup\{c\}\cup\{a,d\}$;
$\{a,b\}\cup\{c,d\}$; $\{b\}\cup\{d\}\cup\{a,c\}$;
$\{a,c\}\cup\{b,d\}$; $\{c\}\cup\{d\}\cup\{a,b\}$;
$\{a,d\}\cup\{b,c\}$;

Для чисел Белла также существует рекуррентное соотношение

$$B_{n+1}=\sum_{k=0}^{n}C_n^kB_k\,. \tag{6}$$

Чтобы его обосновать, зафиксируем в $(n+1)$-элементном множестве A некоторый элемент a. Формируем разбиение множества A, выбирая из множества $A\setminus\{a\}$ содержащий a класс мощности k $(0\le k\le n)$ (C_n^k способов) и разбиваем его дополнение (B_{n-k} способов), откуда и следует (6).

Для B_n существует также красивая *формула Добинского*, выражающая это число в виде суммы бесконечного ряда

$$B_n=\frac{1}{e}\sum_{m=0}^{\infty}\frac{m^n}{m!}, \tag{7}$$

которая будет доказана в разделе 5.

Числом Стирлинга первого рода s_n^m называется число подстановок на n символах, имеющих m независимых циклов. Если подстановка имеет j_1 циклов длины 1, j_2 циклов длины 2, j_3 циклов длины 3 и т. д., где $1j_1+2j_2+\ldots+nj_n=n$, то говорят, что это подстановка типа $(j_1,j_2,\ldots,j_n)$. Записать цикл длиной j_i можно j_i способами, циклически сдвигая элементы цикла. Поэтому, рассуждая как при выводе (1), получаем, что число подстановок типа $(j_1,j_2,\ldots,j_n)$ равно

$$\frac{n!}{\prod_{i=1}^{n}i^{j_i}j_i!}, \tag{8}$$

что для чисел Стирлинга первого рода даёт

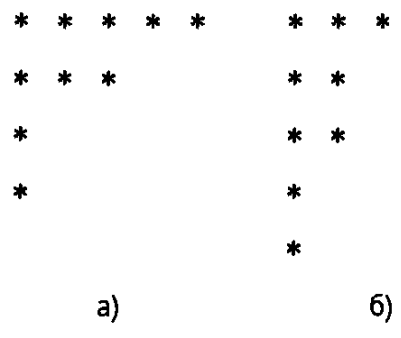
$$s_n^m=\sum_{\substack{j_1,\ldots,j_n\ge 0\\ 1j_1+\ldots+nj_n=n\\ j_1+\ldots+j_n=m}}\frac{n!}{\prod_{i=1}^{n}i^{j_i}j_i!}. \tag{9}$$

Однако для s_n^m также существует удобное рекуррентное соотношение

$$s_n^m = (n-1)s_{n-1}^m + s_{n-1}^{m-1}. \qquad (10)$$

Доказательство (10) аналогично доказательству (4), с той лишь разницей, что убранный из n-элементного множества элемент можно добавить в один из m циклов $(n-1)$-элементного множества $n-1$ способами, так число способов добавление в каждый цикл равно длине цикла. Заметим также, что *числом Стирлинга первого рода* в математической литературе часто называют не s_n^m, а $(-1)^{n-m}s_n^m$.

Разбиения чисел. Если элементы разбиваемого множества неразличимы, то каждое разбиение полностью характеризуется мощностями подмножеств разбиения. Это эквивалентно представлению натурального числа, выражающего мощность разбиваемого множества, в виде суммы натуральных чисел — мощностей подмножеств разбиения. Разумеется, порядок слагаемых здесь не имеет значения, но бóльшие слагаемые принято записывать первыми. Полное число способов разбиения натурального числа n обозначим через $P(n)$, а число способов представления его в виде суммы m слагаемых — через $P_m(n)$. Например, имеем $P(5)=7$: $1+1+1+1+1$, $2+1+1+1$, $3+1+1$, $2+2+1$, $4+1$, $3+2$, 5, причём $P_1(5)=1$, $P_2(5)=2$, $P_3(5)=2$, $P_4(5)=1$, $P_5(5)=1$.



а) б)

Рис. 2

Исследования, связанные с разбиениями чисел, имеют давнюю и богатую историю, начавшуюся с работ Эйлера [20]. Здесь мы лишь слегка коснёмся этого предмета, вернувшись к нему в разделе 5. Наглядным и весьма полезным представлением разбиения является *диаграмма Ферре*, число строк в которой равно числу слагаемых, а число точек в строках — величине слагаемых. Например, разбиение $10=5+3+1+1$ представлено диаграммой Ферре на рис. 2а, а диаграммой на рис. 2б, полученной из 2а заменой строк на столбцы, задаётся разбиение $10=4+2+2+1+1$, которое называется *двойственным* по отношению к исходному.

Максимальное слагаемое в исходном разбиении совпадает по величине с числом слагаемых в двойственном и обратно, а операция сопряжения на множестве разбиений действует как биекция. Отсюда вытекают интересные результаты:

1) *Число разбиений n на m слагаемых равно числу всевозможных разбиений n, в которых наибольшее из слагаемых равно m.*
2) *Число разбиений n на не более чем m слагаемых равно числу всех разбиений n, в которых ни одно из слагаемых не превышает m.*

Вычислять же значения $P_m(n)$ удобно с помощью следующего рекуррентного соотношения

$$P_m(n) = P_m(n-m) + P_{m-1}(n-1). \qquad (11)$$

Для доказательства (11) разобьём множество разбиений числа n на m слагаемых на два класса — класс, в котором все слагаемые больше единицы, и класс, в котором имеются единичные слагаемые. Вычитая по единице из каждого слагаемого разбиения первого класса, можно получить разбиение числа $n-m$ на m слагаемых. Поэтому в первом классе лежит $P_m(n-m)$ разбиений. Убирая единичное слагаемое из разбиения второго класса, можно получить разбиение числа $n-1$ на $m-1$ слагаемых. Поэтому во втором классе $P_{m-1}(n-1)$ разбиений.

Комбинаторные отображения. Многие задачи комбинаторики сводятся к изучению отображений из одного конечного множества в другое. Рассмотрим такие отображения в общем виде. Пусть $f : A \to B$, где $|A| = n$, $|B| = m$. Используя введённые комбинаторные числа, найдём число таких функций. Будем при этом различать случаи, когда все элементы каждого из множеств A или B различимы или неразличимы, а функция f может быть произвольной функцией, а также инъекцией или сюръекцией. Уточним, что в случае неразличимости элементов множества A функции $f_1 : A \to B$ и $f_2 : A \to B$ одинаковы, если некоторой подстановкой элементов множества A одна из фупкций может быть переведена в другую. Аналогичным образом определяется совпадение фупкций в случае неразличимости элементов множества B. Данную задачу можно наглядно проинтерпретировать как размещение n шаров по m урнам, где шары и урны могут быть как различимыми, так и неразличимыми. Результаты подсчётов представлены в табл. 1.

Элементы множества A	*Элементы множества B*	*f — произвольная функция*	*f — инъекция* $(n \le m)$	*f — сюръекция* $(m \le n)$
различимы	*различимы*	m^n	A_m^n	$m!S_n^m$
неразличимы	*различимы*	$\bar{C}_m^n$	C_m^n	$\bar{C}_m^{n-m}$
различимы	*неразличимы*	$\sum_{i=1}^{m} S_n^i$	1	S_n^m
неразличимы	*неразличимы*	$\sum_{i=1}^{m} P_i(n)$	1	$P_m(n)$

Таблица 1

Проверка заполнения данной таблицы будет для читателя хорошим упражнением на комбинаторные числа. Поясним лишь два случая. Если элементы множества A неразличимы, а элементы множества B различимы,

то произвольная функция f определяется набором кратностей принимаемых ею из множества B значений, т. е. является мультимножеством мощности n, определённым на m типах. А в случае, когда элементы множества A различимы, а элементы множества B неразличимы, сюръекция является разбиением n-элементного множества на m классов.

Рассмотрим снова число всевозможных отображений $f: A \to B$, когда элементы обоих множеств различимы. С одной стороны, это число равно m^n. С другой стороны, образ $f(A) \subseteq B$ мощности i можно выбрать C_m^i способами, где $1 \le i \le n$, а для заданного образа мощности i существует $i!S_n^i$ отображений. Это даёт полезное соотношение

$$m^n = \sum_{i=1}^{m} A_m^i S_n^i. \qquad (12)$$

Вопросы для самопроверки

1. Сколькими способами может быть отмечено 5 номеров из 36 при заполнении карточки лотереи?

   а) A_{36}^5; б) C_{36}^5; в) 36^5.

2. Пусть имеется n языков. Сколько нужно издать словарей, чтобы был возможен перевод с любого языка на любой? (англо-русский и русско-английский — это два разных словаря)

   а) C_n^2; б) A_n^2; в) $2n$.

3. У мамы 5 яблок, 7 груш и 3 апельсина. Каждый день, в течение 15 дней, она решила давать сыну по одному фрукту. Сколькими способами это может быть сделано?

   а) C_{15}^3; б) $5 \cdot 7 \cdot 3$; в) $\dfrac{15!}{5!7!3!}$.

4. В распоряжении имеются яблоки, груши и апельсины. Сколькими способами может быть составлен подарочный набор из 5 фруктов?

   а) C_5^3; б) A_5^3; в) $\bar{C}_3^5$.

5. Сколькими способами можно разделить яблоко, грушу, апельсин, мандарин, сливу и айву между тремя мальчиками так, чтобы каждому досталось по 2 фрукта?

   а) C_6^3; б) $\dfrac{6!}{(2!)^3}$; в) 6^3.

6. Пусть в турнире участвуют $2n$ команд. Сколькими способами может быть проведен первый круг, т. е. сколькими способами команды могут быть разбиты на пары?

а) $\frac{(2n)!}{(2!)^n \cdot n!}$; б) $\frac{(2n)!}{2^n}$; в) $\frac{(2n)!}{n!}$.

7. Сколькими способами группу из 20 студентов можно разбить на три подгруппы 7 + 7 + 6 и распределить между тремя преподавателями?

а) $\frac{20!}{6!7!7!}$; б) $\frac{20!}{6!7!7!2!}$; в) $\frac{3 \cdot 20!}{6!7!7!}$.

8. Число Стирлинга второго рода S_n^2 равно

а) 2^n; б) $2^n - 1$; в) $2^{n-1} - 1$.

Ответы: 1 — б, 2 — б, 3 — в, 4 — в, 5 — б, 6 — а, 7 — в, 8 — в.

1.2. Биномиальные коэффициенты

Тут уж буфетчик возмутился. — Это никому не известно и никого не касается, — ответил он. — Ну да, неизвестно, — послышался всё тот же дрянной голос из кабинета, — подумаешь, бином Ньютона!

Михаил Булгаков «Мастер и Маргарита»

Пусть требуется раскрыть выражение $(x_1 + x_2 + \ldots + x_m)^n$, т. е. перемножив n скобок, привести его к виду

$$(x_1 + \ldots + x_m)^n = \underbrace{(x_1 + \ldots + x_m) \cdot \ldots \cdot (x_1 + \ldots + x_m)}_{n \text{ раз}} =$$

$$= \sum_{\substack{i_1, \ldots, i_m \geq 0 \\ i_1 + \ldots + i_m = n}} A(i_1, \ldots, i_m)\, x_1^{i_1} \cdot \ldots \cdot x_m^{i_m}.$$

Числа $A(i_1, \ldots, i_m)$ называются *полиномиальными коэффициентами.* Найдём эти числа. В соответствии с правилами алгебры в процессе раскрытия скобок из каждой скобки выбирается один из символов $\{x_1, \ldots, x_m\}$, и они перемножаются. В результате возникает набор из n сомножителей, выбираемых из множества $\{x_1, \ldots, x_m\}$. И наоборот, каждый такой набор может быть получен как результат подобного выбора. Затем среди полученных таким образом наборов выполняется приведение подобных членов, в результате чего и возникают коэффициенты $A(i_1, \ldots, i_m)$. Поэтому коэффициент $A(i_1, \ldots, i_m)$ равен числу составленных из символов $x_1, \ldots, x_m$

последовательностей длины n, в которых символ x_j используется i_j раз. Как было установлено в предыдущем разделе, число таких последовательностей равно $\frac{n!}{i_1!\cdot i_2!\cdot\ldots\cdot i_m!}$. Это даёт полиномиальную формулу:

$$(x_1+x_2+\ldots+x_m)^n=\sum_{\substack{i_1,\ldots,i_m\geq 0\\ i_1+\ldots+i_m=n}}\frac{n!}{i_1!\cdot i_2!\cdot\ldots\cdot i_m!}\,x_1^{i_1}\cdot x_2^{i_2}\cdot\ldots\cdot x_m^{i_m}\,. \qquad (1)$$

Наиболее часто формула (1) используется при возведении в n-ю степень двучлена. В этом случае она принимает вид

$$(a+b)^n=\sum_{i=0}^{n}\frac{n!}{i!(n-i)!}\,a^i\,b^{n-i}=\sum_{i=0}^{n}C_n^i\,a^i\,b^{n-i} \qquad (2)$$

и называется *биномом Ньютона. Биномиальные коэффициенты* $\frac{n!}{i!(n-i)!}=$ $=C_n^i$ суть величины, фигурировавшие ранее как числа сочетаний.

Из школьного курса математики хорошо известны частные случаи бинома для $n=2$ и $n=3$:

$$(a+b)^2=a^2+2ab+b^2;$$
$$(a+b)^3=a^3+3a^2b+3ab^2+b^3.$$

Биномиальные коэффициенты очень важны и встречаются практически во всех разделах математики. Рассмотрим их важнейшие свойства.

Найдя отношение,

$$\frac{C_n^i}{C_n^{i-1}}=\frac{\frac{n!}{i!(n-i)!}}{\frac{n!}{(i-1)!(n-i+1)!}}=\frac{n!\,(i-1)!\,(n-i+1)!}{n!\,i!\,(n-i)!}=\frac{n-i+1}{i},$$

видим, что оно больше единицы при $i\leq n/2$ и меньше единицы при $m\geq$ $\geq n/2+1$. Поэтому биномиальные коэффициенты C_n^i растут по i от 0 до $[n/2]$ и убывают от $]n/2[$ до n. При чётном n максимальный коэффициент один — $C_n^{n/2}$, при нечётном n максимальных коэффициентов два — $C_n^{[n/2]}$ и $C_n^{]n/2[}$. (Здесь $[a]$ — обозначает максимальное целое, не превосходящее a, а $]a[$ — минимальное целое, не меньшее a.)

Приведём простейшие (и важнейшие) тождества для биномиальных коэффициентов.

$$C_n^m = C_n^{n-m}; \tag{3}$$

$$C_n^m = C_{n-1}^m + C_{n-1}^{m-1}; \tag{4}$$

$$\sum_{i=0}^{n} C_n^i = (1+1)^n = 2^n; \tag{5}$$

$$\sum_{i=0}^{n} (-1)^i C_n^i = (1-1)^n = 0; \tag{6}$$

$$C_{n+m}^k = \sum_{i=0}^{k} C_n^i C_m^{k-i}; \tag{7}$$

$$C_{2n}^n = \sum_{i=0}^{n} \left(C_n^i\right)^2. \tag{8}$$

Все эти тождества могут быть доказаны элементарными алгебраическими выкладками. Поучительны, однако, доказательства, вскрывающие их комбинаторный смысл.

В справедливости тождества (3) можно убедиться, если заметить, что каждое m-элементное подмножество n-элементного множества однозначно определяется своим $(n-m)$-элементным дополнением. Тождество (4) можно получить с помощью следующего рассуждения. Выделим в n-элементном множестве один из элементов. Каждое m-элементное подмножество либо содержит, либо не содержит выделенный элемент. Подмножеств первого типа C_{n-1}^{m-1}, второго — C_{n-1}^m.

Тождества (5) и (6) следуют, как показано, из формулы бинома, причём тождество (5) выражает тот факт, что n - элементное множество имеет 2^n подмножеств. Тождество (7) получим, если рассмотрим разбиение $(n+m)$-элементного множества на n-элементное и m-элементное. Тождество (8) следует из (7), если положить $m = k = n$.

На втором тождестве основан треугольник Паскаля (рис. 1):

1	$(a+b)^0$
1 1	$(a+b)^1$
1 2 1	$(a+b)^2$
1 3 3 1	$(a+b)^3$
1 4 6 4 1	$(a+b)^4$
1 5 10 10 5 1	$(a+b)^5$
. . . . .	

Рис. 1

По бокам треугольника стоят единицы. Каждое число внутри треугольника образуется сложением двух стоящих над ним чисел. В построенном таким образом треугольнике $(n+1)$-я строка является строкой биномиальных коэффициентов C_n^i, $i = 0, 1, \ldots, n$, т. е. коэффициентов разложения $(a+b)^n$. Эта красивая числовая таблица, найденная французским учёным Блезом Паскалем в 1654 году, сыграла важную роль в развитии комбинаторики.

Из биномиальной формулы можно извлечь еще целый ряд соотношений между биномиальными коэффициентами. Перепишем формулу (2) в виде

$$(1+x)^n = \sum_{k=0}^{n} C_n^k x^k . \qquad (9)$$

Продифференцировав (9) по x

$$n(1+x)^{n-1} = \sum_{k=1}^{n} k\, C_n^k x^{k-1}$$

и подставив $x = 1$, получим новое интересное соотношение

$$\sum_{k=1}^{n} k\, C_n^k = n2^{n-1} . \qquad (10)$$

Подставим теперь в (9) вместо x мнимую единицу i

$$(1+i)^n = \sum_{k=0}^{n} C_n^k i^k = (1 - C_n^2 + C_n^4 - C_n^6 + \ldots) + i(C_n^1 - C_n^3 + C_n^5 - C_n^7 + \ldots) .$$

С другой стороны, по формуле Муавра имеем

$$(1+i)^n = \left(\sqrt{2}\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right)\right)^n = 2^{n/2}\left(\cos\frac{n\pi}{4} + i\sin\frac{n\pi}{4}\right) =$$

$$= 2^{n/2}\cos\frac{n\pi}{4} + i2^{n/2}\sin\frac{n\pi}{4} .$$

Приравнивая действительные и мнимые части, получаем ещё два соотношения:

$$1 - C_n^2 + C_n^4 - C_n^6 + \ldots = 2^{n/2}\cos\frac{n\pi}{4} ; \qquad (11)$$

$$C_n^1 - C_n^3 + C_n^5 - C_n^7 + \ldots = 2^{n/2}\sin\frac{n\pi}{4} . \qquad (12)$$

Список получаемых подобным образом соотношений может быть продолжен. Практически, однако, наиболее важными и часто используемыми в самых различных вопросах являются тождества (3)–(6).

С помощью тождества (4) можно получить интересное обобщение тождества (6):

$$\sum_{i=0}^{k}(-1)^i C_n^i = (-1)^k C_{n-1}^k . \qquad (13)$$

В самом деле, используя (4), получаем

$$\sum_{i=0}^{k}(-1)^i C_n^i = \sum_{i=0}^{k}(-1)^i (C_{n-1}^{i-1} + C_{n-1}^i) = C_{n-1}^0 - (C_{n-1}^0 + C_{n-1}^1) + (C_{n-1}^1 + C_{n-1}^2) -$$

$$- (C_{n-1}^2 + C_{n-1}^3) + \ldots + (-1)^{k-1} (C_{n-1}^{k-2} + C_{n-1}^{k-1}) +$$

$$+ (-1)^k (C_{n-1}^{k-1} + C_{n-1}^k) = (-1)^k C_{n-1}^k .$$

А вот ещё одно тождество

$$\sum_{i=m}^{n} (-1)^{i-m} C_n^i C_i^m = \begin{cases} 1, & \text{при } n = m, \\ 0, & \text{при } n > m. \end{cases} \qquad (14)$$

Для его доказательства заметим, прежде всего, что

$$C_n^i C_i^m = \frac{n!}{i!(n-i)!}\frac{i!}{m!(i-m)!} = \frac{n!}{m!(n-m)!}\frac{(n-m)!}{(n-i)!(i-m)!} = C_n^m C_{n-m}^{i-m} .$$

Данное равенство имеет очевидный комбинаторный смысл. Система вложенных множеств $A \subseteq B \subseteq C$ с мощностями, соответственно, m, i и n может быть получена $C_n^i C_i^m$ способами, если сначала из множества C выбирать множество B, а затем из множества B выбирать множество A, и $C_n^m C_{n-m}^{i-m}$ способами, если сначала из множества C выбирать множество A, а затем дополнять его до B.

Теперь доказательство тождества (14) уже не составляет труда:

$$\sum_{i=m}^{n} (-1)^{i-m} C_n^i C_i^m = C_n^m \sum_{i=m}^{n} (-1)^{i-m} C_{n-m}^{i-m} = C_n^m (-1)^m \sum_{j=0}^{n-m} (-1)^j C_{n-m}^j =$$

$$= \begin{cases} 1, \text{ при } n = m, \\ 0, \text{ при } n > m. \end{cases}$$

Вопросы для самопроверки

1. Коэффициент при $x_1^3 x_2^4 x_3^3$ в разложении $(x_1 + x_2 + x_3)^{10}$ равен

а) 10^3; б) C_{10}^3; в) $\frac{10!}{3!4!3!}$.

2. Коэффициент при $a^3b^3c^4$ в разложении $(a+b+c)^2\cdot(a^2+b^2+c^2)^4$ равен

   а) 12; б) 24; в) 18.

3. Коэффициент при t^{17} в разложении $(2+t^4+t^7)^{15}$ равен

   а) $\frac{2^{12}\cdot 15!}{12!2!1!}$; б) 0; в) $\frac{17!}{4!7!}$.

Ответы: 1 — в, 2 — б, 3 — б.

1.3. Формула «включения и исключения»

Чтобы найти мощность объединения двух непересекающихся множеств (рис. 1),

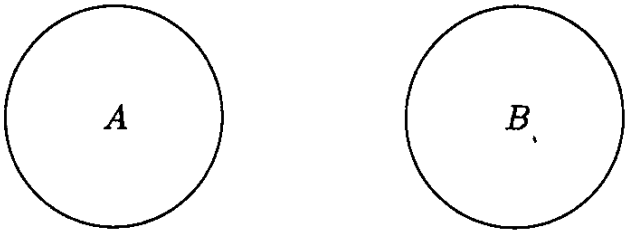


Рис. 1

нужно просто сложить их мощности:

$$|A\cup B| = |A| + |B|.$$

Если же множества пересекаются (рис. 2),

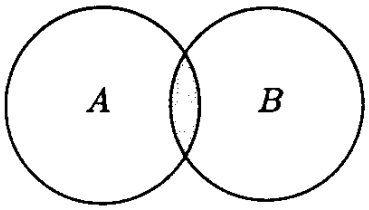


Рис. 2

то при сложении их мощностей каждый элемент объединения, входящий только в одно из множеств, учитывается один раз, но каждый элемент пересечения будет посчитан дважды. Поэтому для правильного ответа необходимо из суммы мощностей вычесть мощность их пересечения:

$$|A\cup B| = |A| + |B| - |A\cap B|.$$

При этом все элементы объединения считаются ровно один раз.

Пусть теперь имеются три множества (рис. 3).

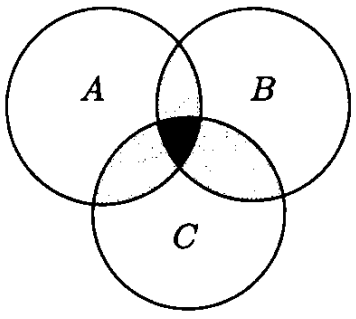


Рис. 3

При сложении мощностей всех трёх множеств каждый элемент, входящий ровно в одно из множеств, учитывается один раз, каждый элемент, входящий ровно в два множества, будет посчитан дважды, а каждый элемент, входящий во все три множества, — трижды. Если из суммы мощностей

вычесть мощности попарных пересечений, то по одному разу будут посчитаны элементы, входящие ровно в одно множество и ровно в два множества, но элементы, входящие во все три множества, не будут посчитаны ни разу. Поэтому для получения правильного ответа необходимо еще прибавить мощность пересечения всех трёх множеств:

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|. \quad (1)$$

В качестве примера применения формулы (1) рассмотрим следующую задачу.

> Каждый студент группы обязан изучать хотя бы один из трёх иностранных языков. Немецкий изучают 6 человек, французский — 7, английский — 8, 3 человека одновременно изучают немецкий и французский, 4 — немецкий и английский, 5 — французский и английский и один человек изучает все три языка. Сколько всего студентов в группе?

$$(6 + 7 + 8) - (3 + 4 + 5) + 1 = 10. \blacktriangleleft$$

Формула, аналогичная (1), справедлива и в общем случае n множеств:

$$|U_1 \cup U_2 \cup \ldots \cup U_n| = \sum_{i=1}^{n} |U_i| - \sum_{1 \le i_1 < i_2 \le n} |U_{i_1} \cap U_{i_2}| + \ldots +$$

$$+ (-1)^{j-1} \sum_{1 \le i_1 < \ldots < i_j \le n} |U_{i_1} \cap U_{i_2} \cap \ldots \cap U_{i_j}| + \ldots + (-1)^{n-1} |U_1 \cap U_2 \cap \ldots \cap U_n|. \quad (2)$$

Эта формула может быть записана компактнее, если ввести следующие обозначения:

$$S_j = \sum_{1 \le i_1 < \ldots < i_j \le n} |U_{i_1} \cap U_{i_2} \cap \ldots \cap U_{i_j}|, \quad j = 1, 2, \ldots, n. \quad (3)$$

Тогда формула (2) примет вид

$$|U_1 \cup U_2 \cup \ldots \cup U_n| = \sum_{j=1}^{n} (-1)^{j-1} S_j. \quad (4)$$

Докажем эту классическую формулу, называемую *«включения и исключения»*. Пусть элемент u входит ровно в k подмножеств $U_{i_1}, \ldots, U_{i_k}$. Тогда в каждой из сумм S_j он будет посчитан C_k^j раз и в правую часть формулы (4) даст вклад, равный $C_k^1 - C_k^2 + \ldots + (-1)^{k-1} C_k^k$. Но из тождества (2.6) для биномиальных коэффициентов следует, что

$$C_k^1 - C_k^2 + \ldots + (-1)^{k-1} C_k^k = 1.$$

Поэтому вклад каждого элемента в правую часть формулы (4) будет равен единице, т. е. правая часть будет равна полному числу элементов, что и доказывает формулу «включения и исключения».

С помощью формулы (4) найдём число костей домино, найденное ранее в разделе 1.1 с помощью формулы для числа сочетаний с повторениями.

Найти с помощью формулы «включения и исключения» число костей домино.

На множестве костей домино можно выделить 7 подмножеств $U_0, U_1, \ldots, U_6$, мощность каждого из которых равна 7. Принадлежность некоторой кости подмножеству U_i характеризуется тем, что хотя бы одно из двух её полей помечено цифрой i. При $i \neq j$ имеем $|U_i \cap U_j| = 1$, так как имеется ровно одна кость, одно поле которой помечено цифрой i, а другое — цифрой j. Поэтому по формуле «включения и исключения» получаем, что число костей равно $|U_0 \cup U_1 \cup \ldots \cup U_6| = 7 \cdot 7 - C_7^2 \cdot 1 = 28$. ◀

В практических задачах часто имеется некоторое множество U и система его подмножеств $U_1, U_2, \ldots, U_n \subseteq U$. Требуется найти число элементов множества U, не принадлежащих ни одному из множеств $U_1, U_2, \ldots, U_n$. В этом случае формула «включения и исключения» выглядит следующим образом

$$|U \setminus (U_1 \cup U_2 \cup \ldots \cup U_n| = |U| - \sum_{i=1}^{n} |U_i| + \sum_{1 \le i_1 < i_2 \le n} |U_{i_1} \cap U_{i_2}| + \ldots +$$

$$+ (-1)^j \sum_{1 \le i_1 < \ldots < i_j \le n} |U_{i_1} \cap U_{i_2} \cap \ldots \cap U_{i_j}| + \ldots + (-1)^n |U_1 \cap U_2 \cap \ldots \cap U_n|. \quad (5)$$

Воспользовавшись обозначениями (3) и положив $S_0 = |U|$, формулу (5) можно представить в виде

$$|U \setminus (U_1 \cup U_2 \cup \ldots \cup U_n| = \sum_{j=0}^{n} (-1)^j S_j. \quad (6)$$

Рассмотрим примеры.

Найти число натуральных чисел, не превосходящих 100 и не делящихся ни на одно из чисел 3, 5, 7.

Число чисел, делящихся на 3, равно $[100/3] = 33$; на 5 — $[100/5] = 20$; на 7 — $[100/7] = 14$. Число чисел, делящихся на 3 и 5, равно $[100/15] = 6$; на 3 и 7 — $[100/21] = 4$, на 5 и 7 — $[100/35] = 2$. Число чисел, делящихся на все три числа 3,

5 и 7, равно $[100/105]=0$. Поэтому искомое число равно $100-(33+20+14)+$ $+(6+4+2)-0=45$.◀

А теперь рассмотрим пример посложнее.

Найти число целочисленных решений системы

$$\begin{cases} x_1+x_2+x_3=40, \\ 4\le x_1\le 15, \\ 9\le x_2\le 18, \\ 5\le x_3\le 16. \end{cases}$$

Формула «включения и исключения» оказывается полезной и здесь. Введём новые переменные $y_1=x_1-4$, $y_2=x_2-9$, $y_3=x_3-5$. Система перепишется в виде

$$\begin{cases} y_1+y_2+y_3=22, \\ 0\le y_1\le 11, \\ 0\le y_2\le 9, \\ 0\le y_3\le 11. \end{cases}$$

Пусть U — множество решений системы

$$\begin{cases} y_1+y_2+y_3=22, \\ y_1,y_2,y_3\ge 0, \end{cases}$$

U_1 — множество решений системы

$$\begin{cases} y_1+y_2+y_3=22, \\ y_1\ge 12, \\ y_2,y_3\ge 0, \end{cases}$$

U_2 — множество решений системы

$$\begin{cases} y_1+y_2+y_3=22, \\ y_2\ge 10, \\ y_1,y_1\ge 0, \end{cases}$$

U_3 — множество решений системы

$$\begin{cases} y_1+y_2+y_3=22, \\ y_3\ge 12, \\ y_1,y_2\ge 0. \end{cases}$$

Для мощности множества U согласно разделу 1 имеем

$$|U|=\bar{C}_3^{22}=C_{24}^{22}=276.$$

Чтобы найти мощность множества U_1, достаточно в соответствующей системе сделать замену $z_1 = y_1 - 12$. Это дает

$$|U_1| = \bar{C}_3^{10} = C_{12}^{10} = 66.$$

Аналогично,

$$|U_2| = \bar{C}_3^{12} = C_{14}^{12} = 91, \; |U_3| = 66.$$

Далее, легко видеть, что

$$|U_1 \cap U_3| = 0, \; |U_1 \cap U_2| = 1, \; |U_2 \cap U_3| = 1.$$

Поэтому в соответствии с формулой «включения и исключения» число решений исходной системы равно

$$|U \setminus (U_1 \cup U_2 \cup U_3)| = |U| - |U_1| - |U_2| - |U_3| + |U_1 \cap U_2| +$$
$$+ |U_2 \cap U_3| + |U_1 \cap U_3| - |U_1 \cap U_2 \cap U_3| =$$
$$= 276 - 66 - 91 - 66 + 1 + 1 + 0 - 0 = 55. \quad \blacktriangleleft$$

В качестве ещё одного примера рассмотрим классическую задачу о беспорядках.

> Найти число перестановок чисел $\{1, 2, \ldots, n\}$, в которых никакое число i не стоит на i-м месте.

Всего перестановок $|U| = n!$. Перестановок, в которых число i стоит на i-м месте, $|U_i| = (n-1)!$ Перестановок, в которых два различных числа i и j стоят на своих местах, $|U_i \cap U_j| = (n-2)!$ и т. д. Таким образом, используя обозначения (3), имеем

$$S_j = C_n^j (n-j)! = \frac{n!}{j!},$$

и по формуле (6) получаем

$$|U \setminus (U_1 \cup U_2 \cup \ldots \cup U_n| = \frac{n!}{2!} - \frac{n!}{3!} + \frac{n!}{4!} - \frac{n!}{5!} + \ldots + (-1)^n$$
$$= n! \left(\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \ldots + \frac{(-1)^n}{n!} \right). \quad \blacktriangleleft$$

Отметим, что выражение в скобках с ростом n стремится к e^{-1}, так как

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots.$$

А теперь сходным приёмом докажем формулу Стирлинга (1.5). Разбиение n-элементного множества на m подмножеств можно рассматривать как размещение n различных шаров по m одинаковым урнам, причём в каждую урну должен попасть хотя бы один шар. В табл. 1.1 соответствующая функция f является сюръекцией.

Найдём число размещений в случае различных урн, т. е. вычислим $m!S_n^m$. Полное число размещений, включая и размещения с пустыми урнами, есть m^n, а число размещений, в которых некоторая фиксированная урна остаётся пустой, — $(m-1)^n$. Рассуждая как и в предыдущей задаче, получаем

$$m!S_n^m = \sum_{i=0}^{m-1}(-1)^i C_m^i (m-i)^n = \sum_{i=1}^{m}(-1)^{m-i} C_m^i i^n ,$$

откуда и следует (1.5).

При применении формулы «включения и исключения» вычисление всех входящих в (2) сумм может оказаться затруднительным. Что будет, если ограничиться первыми $l < n$ суммами, т. е. рассматривать пересечения не более чем l подмножеств? Ясно, что справедливо неравенство

$$\sum_{i=1}^{n}|U_i| - \sum_{1\le i_1<i_2\le n}|U_{i_1}\cap U_{i_2}| \le |U_1\cup U_2\cup\ldots\cup U_n| \le \sum_{i=1}^{n}|U_i| .$$

Аналогичные неравенства, называемые *неравенствами Бонферрони*, справедливы и в общем случае

$$\sum_{j=1}^{2m}(-1)^{j-1}S_j \le |U_1\cup U_2\cup\ldots\cup U_n| \le \sum_{j=1}^{2m+1}(-1)^{j-1}S_j . \quad (7)$$

То есть при сохранении лишь первых $l < n$ сумм искомая мощность объединения множеств вычисляется с недостатком, если первая отбрасываемая сумма имеет положительный знак, и с избытком — если отрицательный. При этом абсолютная величина погрешности, очевидно, не превосходит абсолютной величины первого отбрасываемого члена.

Чтобы это доказать, снова рассмотрим элемент u, входящий в k подмножеств $U_{i_1}, \ldots, U_{i_k}$. Вклад, даваемый этим элементом в усечённую до l членов правую часть формулы (4), равен

$$C_k^1 - C_k^2 + \ldots + (-1)^{l-1}C_k^l = \sum_{i=1}^{l}(-1)^{i-1}C_k^i + \sum_{i=l+1}^{k}(-1)^{i-1}C_k^i - \sum_{i=l+1}^{k}(-1)^{i-1}C_k^i =$$

$$= \sum_{i=1}^{k}(-1)^{i-1}C_k^i - \sum_{i=l+1}^{k}(-1)^{i-1}C_k^i = \sum_{i=1}^{k}(-1)^{i-1}C_k^i +$$

$$+ (-1)^{k+1}\sum_{i=0}^{k-l-1}(-1)^i C_k^i = 1 + (-1)^l C_{k-1}^{k-l-1}.$$

В последнем равенстве было использовано тождество (2.13). Таким образом, этот вклад оказывается большим или меньшим единицы в зави-

симости от знака первой отбрасываемой суммы. В рассмотренной выше задаче о беспорядках число полностью лишённых порядка перестановок будет вычисляться с недостатком или с избытком в зависимости от знака первого отбрасываемого члена.

Подмножества $U_1, U_2, \ldots, U_n$, как правило, задаются предикативно, т. е. принадлежность элемента данному подмножеству определяется наличием у него определённого свойства. Возникает задача нахождения числа элементов, обладающих ровно l свойствами и не менее чем l свойствами.

Число элементов, принадлежащих ровно l подмножествам, равно

$$\sum_{j=l}^{n} (-1)^{j-l} C_j^l S_j . \qquad (8)$$

Чтобы установить это, достаточно заметить, что элемент u, принадлежащий $k \geq l$ подмножествам, в каждой сумме S_j учитывается C_k^j раз и в правую часть формулы даёт вклад, равный

$$\sum_{j=l}^{n} (-1)^{j-l} C_j^l C_k^j = \begin{cases} 1, & k = l, \\ 0, & k > l, \end{cases}$$

в соответствии с (2.14).

Используя (8), для числа элементов, принадлежащих не менее чем l подмножествам, получаем

$$\sum_{i=l}^{n}\sum_{j=i}^{n} (-1)^{j-i} C_j^i S_j = \sum_{j=l}^{n} S_j \sum_{i=l}^{n} (-1)^{j-i} C_j^i = \sum_{j=l}^{n} S_j \sum_{i=l}^{j} (-1)^{j-i} C_j^i .$$

В двойной сумме здесь был изменён порядок суммирования, что часто используется в дискретной математике. Для обоснования этого приёма достаточно заметить, что для нахождения суммы чисел в прямоугольной таблице можно просуммировать их в каждой строке и затем сложить полученные суммы, а можно сначала суммировать по столбцам.

Используя (2.3) и (2.13), далее имеем

$$\sum_{j=l}^{n} S_j \sum_{i=l}^{j} (-1)^{j-i} C_j^i = \sum_{j=l}^{n} S_j (-1)^{j-l} C_{j-1}^{j-l} = \sum_{j=l}^{n} (-1)^{j-l} C_{j-1}^{l-1} S_j .$$

Итак, число элементов, принадлежащих не менее чем l подмножествам, есть

$$\sum_{j=l}^{n} (-1)^{j-l} C_{j-1}^{l-1} S_j . \qquad (9)$$

Применим формулы (8) и (9) для нахождения числа перестановок n символов, оставляющих на месте 1) l символов и 2) не менее чем l символов. Считая, что перестановка принадлежит множеству U_i, если она оставляет на месте символ i, получаем, что число перестановок, оставляющих на месте ровно l символов, равно

$$\sum_{j=l}^{n}(-1)^{j-l}C_j^l S_j=\sum_{j=l}^{n}(-1)^{j-l}C_j^l C_n^j (n-j)!=$$

$$=\sum_{j=l}^{n}(-1)^{j-l}\frac{j!}{l!(j-l)!}\frac{n!}{j!(n-j)!}(n-j)!=\frac{n!}{l!}\sum_{j=l}^{n}\frac{(-1)^{j-l}}{(j-l)!}=\frac{n!}{l!}\sum_{i=0}^{n-l}\frac{(-1)^i}{i!}.$$

А число перестановок, оставляющих на месте не менее l символов, есть

$$\sum_{j=l}^{n}(-1)^{j-l}C_{j-1}^{l-1}S_j=\sum_{j=l}^{n}(-1)^{j-l}C_{j-1}^{l-1}C_n^j(n-j)!=$$

$$=\sum_{j=l}^{n}(-1)^{j-l}\frac{(j-1)!}{(l-1)!(j-l)!}\frac{n!}{j!(n-j)!}(n-j)!=$$

$$=\frac{n!}{(l-1)!}\sum_{j=l}^{n}\frac{(-1)^{j-l}}{j(j-l)!}=\frac{n!}{(l-1)!}\sum_{i=0}^{n-l}\frac{(-1)^i}{(l+i)\,i!}.$$

Вопросы для самопроверки

1. В группе 5 студентов не занимаются ни в одной спортивной секции, 10 студентов занимаются ровно в одной из спортивных секций, 6 студентов ходят в две секции и один студент занимается в трёх секциях. Сколько всего студентов в группе?

   а) 22; б) 20; в) 25.

2. Всего в группе 25 студентов. Из них в бассейн ходят 10 человек, в гимнастический зал — 8 человек, в волейбольную секцию — 6 человек. При этом 4 человека ходят одновременно в бассейн и на гимнастику, 3 человека — в бассейн и на волейбол и 2 человека — на гимнастику и на волейбол. Один человек ходит во все три секции. Сколько студентов группы не занимается в спортивных секциях?

   а) 12; б) 9; в) 11.

3. Сколько натуральных чисел, не превосходящих 100, не делятся ни на 2, ни на 3?

   а) 30; б) 33; в) 28.

Ответы: 1 — а, 2 — б, 3 — б.

1.4. Приложения к теории вероятностей

Развитые в предыдущих разделах методы подсчёта позволяют решать задачи теории вероятностей с конечным множеством равновероятных исходов. Задачи подобного типа особенно часто возникают в теории азартных игр. Это возвращает нас к истокам теории вероятностей, которая возникла из анализа шансов при игре в кости.

Для решения подобных задач достаточно классического определения Лапласа, согласно которому вероятность события A есть отношение числа элементарных исходов, благоприятствующих событию A, к полному числу возможных элементарных исходов:

$$\mathrm{P}(A)=\frac{\textit{число элементарных исходов, благоприятствующих } A}{\textit{полное число возможных элементарных исходов}}.$$

Большое число подобных задач теории вероятностей может быть сформулировано в терминах следующей модели.

Из урны, содержащей N шаров, R — красных и $N-R$ — синих, вынимается n шаров. Какова вероятность, что среди них будет r красных и $n-r$ синих шаров?

Любое n-элементное подмножество N-элементного множества шаров будем считать возможным элементарным исходом. Поэтому полное число элементарных исходов равно C_N^n. Так как r красных шаров выбирается из R-элементного множества, а $n-r$ синих — из $N-R$, то выборок, содержащих r красных и $n-r$ синих шаров, в соответствии с правилом произведения имеется $C_N^r \cdot C_{N-R}^{n-r}$. Поэтому искомая вероятность равна

$$\mathrm{P}(A)=\frac{C_R^r \cdot C_{N-R}^{n-r}}{C_N^n}. \blacktriangleleft$$

Рассмотрим ряд примеров на применение данной формулы.

В коробке находятся 10 деталей, из которых 6 стандартных и 4 дефектных. Найти вероятность, что среди наугад взятых трёх деталей будет по крайней мере две стандартных детали?

Стандартные детали будем считать красными шарами, а дефектные — синими. Красных шаров должно быть два или три. Поэтому искомая вероятность есть

$$\mathrm{P}(A)=\frac{C_6^2 \cdot C_4^1 + C_6^3}{C_{10}^3}=\frac{2}{3}. \blacktriangleleft$$

В лотерее из 49 номеров 6 являются выигрышными. Какова вероятность, что среди 6 отмеченных номеров окажется ровно 4 выигрышных?

Здесь выигрышные номера играют роль красных шаров, а невыигрышные — синих. Искомая вероятность есть

$$P(A)=\frac{C_6^4\cdot C_{43}^2}{C_{49}^6}.\ \blacktriangleleft$$

Рассмотрим теперь другой, менее очевидный пример, при решении которого также можно воспользоваться урновой моделью.

> Два одинаково метких стрелка Пётр и Иван состязаются в стрельбе по мишени. Условия состязания таковы, что Пётр делает 5 выстрелов, а Иван — 10. Победа присуждается Ивану, если ему принадлежат оба из двух ближайших к центру мишени выстрелов, и Петру — если среди этих двух есть хотя бы один его выстрел. Найти вероятность победы для каждого из участников.

Здесь роль шаров играют выстрелы, причём выстрелы Петра можно уподобить красным шарам, а выстрелы Ивана — синим. Из 15 выстрелов выбираются 2 определяющих исход соревнования ближайших к центру мишени выстрела. Так как стрелки одинаково меткие, то это соответствует случайному вынимаиию двух шаров из 15. Победа Ивана соответствует тому, что оба вынутых шара оказываются синими:

$$P\left\{\begin{matrix}победа\\Ивана\end{matrix}\right\}=\frac{C_{10}^2}{C_{15}^2}=\frac{3}{7}.$$

Победа присуждается Петру, если оба вынутых шара оказываются красными или один из них красный, а другой синий:

$$P\left\{\begin{matrix}победа\\Петра\end{matrix}\right\}=\frac{C_5^2+C_5^1\cdot C_{10}^1}{C_{15}^2}=\frac{4}{7}.$$

Так как состязание заканчивается победой одного из участников, то сумма двух вероятностей равна, разумеется, единице. ◀

Если из n-элементного множества берётся случайная m-элементная выборка, то полное число элсментарных исходов равно A_n^m, если порядок выбираемых элементов принимается во внимание, и C_n^m — если порядок безразличен. В качестве примеров рассмотрим случайные извлечения карт из стандартной 36-карточной колоды.

> Из перетасованной колоды последовательно тянутся 3 карты. Какова вероятность, что эти 3 карты будут семёрка, дама, туз в заданном порядке?

$$P(A)=\frac{4^3}{A_{36}^3},$$

так как в колоде имеются четыре семёрки, четыре дамы и четыре туза. ◀

Из перетасованной колоды извлекаются 4 карты. Какова вероятность, что будут извлечены 2 короля и 2 дамы?

$$P(A)=\frac{C_4^2\cdot C_4^2}{C_{36}^4},$$

так как 2 короля выбираются из 4 королей, а 2 дамы — из 4 дам. ◄

Из перетасованной колоды извлекаются 3 карты. Какова вероятность, что все они будут различных достоинств?

$$P(A)=\frac{C_9^3\cdot 4^3}{C_{36}^3},$$

так как три достоинства из девяти могут быть выбраны C_9^3 способами, и имеется по 4 карты каждого достоинства в колоде. ◄

Рассмотрим теперь популярную карточную игру в «подкидного дурака», в которой каждый из играющих получает при сдаче по 6 карт, а открываемая карта указывает козырную масть.

Какова вероятность при игре в «подкидного дурака» получить при сдаче 2 туза?

$$P(A)=\frac{C_4^2\cdot C_{32}^4}{C_{36}^6},$$

так как 2 туза дополняются 4 картами, выбранными из 32 карт, среди которых нет тузов. Решение данной задачи может быть выражено и в терминах урновой модели. ◄

Какова вероятность при игре в «подкидного дурака» получить при сдаче 2 козыря?

В этой задаче в элементарный исход наряду с множеством из 6 получаемых при сдаче карт должна быть отдельно включена также и карта, указывающая козырную масть. Если 6 карт сдачи выбрано, то существует 30 возможностей выбрать из оставшихся тридцати карт указывающую козырную масть карту. Поэтому в соответствии с правилом произведения полное число элементарных исходов равно $30\cdot C_{36}^6$. Чтобы найти число элементарных исходов, благоприятствующих данному событию, будем считать, что сначала выбирается козырная масть (4 возможности), затем в ней выбирается указывающая козырную масть открываемая карта (9 возможностей) и, наконец, из оставшихся 8 карт масти выбираются 2 козыря (C_8^2 возможностей), к которым добавляются 4 некозырные карты (C_{27}^4 возможностей). В соответствии с правилом произведения число благоприятных элементарных исходов равно $4\cdot 9\cdot C_8^2\cdot C_{27}^4$, и искомая вероятность

$$\mathrm{P}(A)=\frac{4\cdot 9\cdot C_8^2\cdot C_{27}^4}{30\cdot C_{36}^6}.\blacktriangleleft$$

Какова вероятность, что среди 6 полученных при сдаче карт будут представлены все масти?

Здесь число благоприятствующих исходов может быть найдено с помощью формулы включения и исключения. Пусть U — множество всех возможных сдач; U_1 — множество сдач, не содержащих пик; U_2 — множество сдач, не содержащих треф; U_3 — множество сдач, не содержащих бубен; U_4 — множество сдач, не содержащих червей. Имеем:

$$|U|=C_{36}^6,\quad |U_1|=|U_2|=|U_3|=|U_4|=C_{27}^6,$$
$$|U_i\cap U_j|=C_{18}^6,\quad |U_i\cap U_j\cap U_k|=C_9^6.$$

Отсюда получаем, что искомая вероятность равна

$$\mathrm{P}(A)=\frac{C_{36}^6-C_4^1C_{27}^6+C_4^2C_{18}^6-C_4^3C_9^6}{C_{36}^6}.\blacktriangleleft$$

Подчеркнём, что вероятности всех рассмотренных событий при игре в «подкидного дурака» определялись индивидуальным набором полученных карт и не зависели от числа играющих.

Много содержательных вероятностных интерпретаций можно дать рассмотренной в предыдущем разделе задаче о беспорядках. Например, n мужчин сдают свои шляпы в гардероб и получают их обратно случайным образом. Какова вероятность, что ни на ком не будет одета его собственная шляпа? В другой интерпретации, n супружеских пар пришли на танцевальный вечер, где танцевальные пары составляются по жребию. Какова вероятность, что ни одна супружеская пара не будет танцевать вместе? В этой задаче полное число элементарных исходов равно числу перестановок, т. е. $n!$, поэтому искомая вероятность есть

$$\mathrm{P}(A)=\frac{1}{2!}-\frac{1}{3!}+\ldots+\frac{(-1)^n}{n!},$$

и с ростом n она стремится к e^{-1}.

Рассмотрим теперь более общую по сравнению с классической вероятностью модель, которая называется *конечным вероятностным пространством* и задаётся следующим образом. Пусть имеется конечное множество элементарных исходов $\Omega=\{\omega_1,\ldots,\omega_n\}$, которые называют также точками вероятностного пространства, и пусть каждому $\omega_i\in\Omega$ приписано число

$p(\omega_i)$, где $0 \le p(\omega_i) \le 1$, причём $\sum_{i=1}^{n} p(\omega_i) = 1$. Тогда *конечное вероятностное пространство* задано, если любое подмножество $A \subseteq \Omega$ считается *событием*, а его *вероятность* $\mathrm{P}(A)$ определяется по формуле

$$\mathrm{P}(A) = \sum_{\omega_i \in A} p(\omega_i).$$

Полное число событий равно, таким образом, 2^n. Сюда входит и пустое множество $\varnothing$, вероятность которого равна нулю и которое называется *невозможным событием*. Рассмотренное ранее классическое определение вероятности является, как легко видеть, частным случаем этой модели, когда вероятности элементарных исходов одинаковы и равны $p(\omega_i) = 1/n$, $i = 1, 2, \ldots, n$.

Так как события в конечном вероятностном пространстве определены как подмножества универсального множества Ω, на множестве событий определены теоретико-множественные операции объединения, пересечения и дополнения $A \cup B$, $A \cap B$, $\overline{A}$, или, как говорят, задана *алгебра событий*. Формула «включения и исключения», сохраняется и для вероятностей. В частности,

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B),$$

$$\mathrm{P}(A \cup B \cup C) = \mathrm{P}(A) + \mathrm{P}(B) + \mathrm{P}(C) - \mathrm{P}(A \cap B) - \mathrm{P}(A \cap B) - \mathrm{P}(A \cap B) + \mathrm{P}(A \cap B \cap C)$$

и т. д. Справедливы также неравенства Бонферрони. В частности,

$$\mathrm{P}(A \cup B) \le \mathrm{P}(A) + \mathrm{P}(B),$$

причём если $A \cap B = \varnothing$, т. е. события A и B не пересекаются (несовместны), то здесь имеет место равенство:

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B).$$

Эти соотношения обобщаются на произвольное число событий

$$\mathrm{P}(A_1 \cup \ldots \cup A_k) \le \mathrm{P}(A_1) + \ldots + \mathrm{P}(A_k) \qquad (1)$$

и

$$\mathrm{P}(A_1 \cup \ldots \cup A_k) = \mathrm{P}(A_1) + \ldots + \mathrm{P}(A_k),$$

если события $A_1, \ldots, A_k$ попарно несовместны.

Схема Бернулли. Пусть теперь некоторый опыт со случайным исходом повторяется несколько раз подряд при неизменных условиях. Свяжем с этим опытом некоторое событие A, и будем называть его успехом, а противоположное событие $\overline{A}$ — неудачей. Пусть $\mathrm{P}(A) = p$, $\mathrm{P}(\overline{A}) = 1 - p = q$. Тогда

можно поставить вопрос о нахождении вероятности $P_n(k)$ того, что в серии из n испытаний произойдёт k успехов.

Подобная модель повторных независимых испытаний называется схемой Бернулли, по имени выдающегося швейцарского математика Якоба Бернулли (1654–1705), рассмотревшего эту схему и доказавшего *закон больших чисел*, согласно которому с ростом числа испытаний n частота успехов k/n будет приближаться к вероятности p появления успеха в отдельном испытании. Согласно закону больших чисел, например, при многократном подбрасывании симметричной монеты ($p = q = 1/2$) доля выпадений орла от общего числа подбрасываний будет близка к $1/2$.

В схеме Бернулли элементарными исходами являются последовательности длины n, состоящие из успехов и неудач. Для заданной последовательности, содержащей k успехов и $n-k$ неудач, её вероятность, вычисляемая как произведение вероятностей n независимых событий, равна $p^k q^{n-k}$. Но существует C_n^k таких последовательностей с k успехами и $n-k$ неудачами. Поэтому вероятность того, что в серии из n испытаний произойдёт k успехов равна

$$P_n(k) = C_n^k p^k q^{n-k}. \qquad (2)$$

Вычисляя с помощью формулы (2) вероятность того, что при десятикратном подбрасывании монеты орёл выпадет пять раз, получаем

$$P_{10}(5) = C_{10}^5 \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^5 = \frac{10!}{5!5!} \cdot \frac{1}{1024} = \frac{63}{256} \approx \frac{1}{4}.$$

Вероятность же того, что число успехов в серии из n испытаний Бернулли будет лежать в интервале от k_1 до k_2, с помощью формулы (2) можно найти как

$$P_n\{k_1 \le k \le k_2\} = \sum_{k=k_1}^{k_2} C_n^k p^k q^{n-k}. \qquad (3)$$

Проанализировав формулу (3), Якоб Бернулли показал, что для любого сколь угодно малого ε

$$P_n\{n(p-\varepsilon) \le k \le n(p+\varepsilon)\} \to 1, \quad n \to \infty,$$

что и явилось первым доказательством закона больших чисел. В Добавлении 2 заинтересованный читатель может ознакомиться с современным компактным доказательством этого закона, основанном на неравенстве Чебышёва.

В заключении заметим, что введённое здесь конечным вероятностным пространством не исчерпываются вероятностные модели. В частности, рассматривавшиеся во Вводной главе геометрические вероятности имеют кон-

тинуальное множество элементарных исходов, и поэтому не могут быть описаны в рамках модели конечного вероятностного пространства. Однако этой модели, как правило, оказывается достаточно при применении теории вероятностей к задачам дискретной математики (см. Добавление 2).

Вопросы для самопроверки

1. Из конфетницы, содержащей 4 шоколадных конфеты и 8 карамелей, наугад берутся 2 конфеты. Какова вероятность, что обе они шоколадные?

   а) $\frac{1}{C_4^2}$; б) $\frac{C_4^2}{C_8^2}$; в) $\frac{C_4^2}{C_{12}^2}$.

2. Из перетасованной 36-карточной колоды берутся 3 карты. Какова вероятность, что среди них не будет тузов?

   а) $\frac{C_4^0}{C_{36}^3}$; б) $\frac{C_{32}^3}{C_{36}^3}$; в) $\frac{C_4^3}{C_{36}^3}$.

3. Колода из 36 карт случайным образом делится пополам. Какова вероятность, что в каждой половине будет по 2 туза?

   а) $\frac{C_4^2}{2^4}$; б) $\frac{C_4^2}{C_{36}^{18}}$; в) $\frac{C_4^2 \cdot C_{32}^{16}}{C_{36}^{18}}$.

4. Найти вероятность того, что среди 6 карт, полученных при сдаче в игре в «подкидного дурака», не будет ни одного козыря.

   а) $\frac{4 \cdot 9 \cdot C_{27}^6}{30 \cdot C_{36}^6}$; б) $\frac{C_{27}^6}{C_{36}^6}$; в) $\frac{4 \cdot C_{27}^6}{C_{36}^6}$.

5. Из 8 букв разрезной азбуки составляется слово «институт». Затем составляющие слово карточки перемешиваются и снова собираются в произвольном порядке. Какова вероятность, что снова получится слово «институт»?

   а) $\frac{1}{2^8}$; б) $\frac{1}{8!}$; в) $\frac{2! \cdot 3!}{8!}$.

Ответы: 1 — в, 2 — б, 3 — в, 4 — а, 5 — в.

1.5. Производящие функции и рекуррентные соотношения

В результате выполнения алгебраических преобразований могут осуществляться определенные комбинаторные подсчеты. С этим мы уже встретились в (2.2) при выводе биномиальной формулы, когда комбинаторные объекты — числа сочетаний — появились в новом качестве биномиаль-

ных коэффициентов при выполнении алгебраической операции возведения в степень. Рассмотрим еще примеры, иллюстрирующие эту связь между алгеброй и комбинаторикой.

> Пусть в урне находятся 3 красных, 4 синих и 2 зелёных шара. Сколькими способами могут быть извлечены 6 шаров, если шары одного цвета считать неразличимыми?

Для ответа на этот вопрос достаточно по обычным алгебраическим правилам раскрыть скобки в выражении

$$(1+x+x^2+x^3)(1+x+x^2+x^3+x^4)(1+x+x^2),$$

привести подобные члены и взять коэффициент при x^6. Здесь первая скобка соответствует красным шарам, вторая — синим и третья — зелёным. Степени буквы x в каждой скобке выражают количество шаров данного цвета, которые могут быть взяты. При перемножении скобок, показатели степеней складываются, поэтому x^6 возникнет столько раз, сколько существует способов выбора 6 шаров. Это рассуждение справедливо не только для x^6, но и для любой другой степени x. Выполнив умножение, получим многочлен

$$1+3x+6x^2+9x^3+11x^4+11x^5+9x^6+6x^7+3x^8+x^9,$$

который называется *производящей функцией* для числа способов выбора, так как коэффициент при x^k равен числу способов выбора k шаров. В частности, 6 шаров могут быть извлечены 9 способамн. ◀

Несколько видоизменим задачу.

> Пусть теперь требуется найти, сколькими способами можно извлечь из той же урны 6 шаров так, чтобы среди них было нечётное число красных шаров, чётное число синих и хотя бы один зелёный шар

В этом случае также легко выписать производящую функцию, дающую ключ к решению задачи:

$$(x+x^3)(1+x^2+x^4)(x+x^2)=x^2+x^3+2x^4+2x^5+2x^6+2x^7+x^8+x^9.$$

Теперь 6 шаров могут быть извлечены лишь двумя способами. ◀

Другой пример.

> Пусть имеются четыре монеты достоинством 1 рубль, три монеты — 2 рубля, две монеты — 5 рублей и одна монета — 10 рублей. Сколькими способами можно набрать денежную сумму в размере 10 рублей?

Рассмотрим выражение

$$(1+x+x^2+x^3+x^4)(1+x^2+x^4+x^6)(1+x^5+x^{10})(1+x^{10}),$$

в котором каждая скобка соответствует одному из номиналов. Показатели степени, в которые возводится буква x в скобке, равны произведению номинала монеты на число монет данного номинала, которое может быть включено в набираемую де-

нежную сумму. Это число изменяется от нуля до количества имеющихся в распоряжении монет данного номинала. Так как при умножении показатели степени складываются, то ответ на вопрос задачи даст после перемножения скобок коэффициент при x^{10}. Выполнив умножение, получим:

$$1+x+2x^2+3x^3+3x^4+3x^5+4x^6+4x^7+4x^8+4x^9+5x^{10}+5x^{11}+6x^{12}+$$
$$+7x^{13}+7x^{14}+6x^{15}+7x^{16}+6x^{17}+6x^{18}+5x^{19}+5x^{20}+4x^{21}+4x^{22}+$$
$$+4x^{23}+4x^{24}+3x^{25}+3x^{26}+2x^{27}+2x^{28}+x^{29}+x^{30}.$$

Таким образом, набрать 10 рублей можно пятью способами. Вот они:

$$\{10\},\{5+5\},\{5+2+1+1+1\},\{5+2+2+1\},\{2+2+2+1+1+1+1\}.\blacktriangleleft$$

Конечно, проще было бы решить задачу, просто указав 5 возможных вариантов. Однако полученный в результате перемножения скобок многочлен позволяет для любого n дать ответ на вопрос, сколькими способами может быть получена денежная сумма в размере n рублей. Он является производящей функцией числа способов набора денежных сумм в данной задаче.

Метод производящих функций берёт своё начало в XVIII веке в работах Муавра, Эйлера, Лагранжа и Лапласа, предложившего и сам термин «производящая функция». Идея метода состоит в том, чтобы сопоставить последовательности $\{a_n\}=a_0, a_1, a_2,...$ решений задачи степенной ряд $A(x)=a_0+$ $+a_1x+a_2x^2+...+a_nx^n+...$, который называется *производящей функцией последовательности*. Аналитические методы работы с функцией $A(x)$ часто оказывается проще и эффективнее комбинаторных методов работы с членами последовательности. Если ряд $A(x)$ сходится в некоторой окрестности нуля, то его можно почленно дифференцировать. Поэтому знание производящей функции позволяет восстановить исходную последовательность:

$$a_n=\frac{1}{n!}\cdot\frac{d^nA(x)}{dx^n}\bigg|_{x=0}.$$

Вот как образно характеризует метод производящих известный математик и педагог Дьердь Пойа:

> Производящая функция является устройством, отчасти напоминающим мешок. Вместо того чтобы нести отдельно много маленьких предметов, что могло бы оказаться затруднительным, мы кладём их в мешок, и тогда нам нужно нести лишь один предмет, мешок. Совершенно таким же образом, вместо того чтобы иметь дело с каждым членом последовательности $a_0, a_1, a_2, ..., a_n, ...$ в отдельности, мы ставим их в степенной ряд $\sum_n a_nx^n$ и тогда нам нужно иметь дело лишь с одним математическим объектом, степенным рядом.
>
> *Дьердь Пойа «Математика и правдоподобные рассуждения»*

Полезно знать производящие функции для простейших последовательностей. Для последовательности 1, 1, ... , 1, ... , все члены которой равны единице, производящая функция есть

$$A(x)=1+x+x^2+\ldots+x^n+\ldots=\frac{1}{1-x}. \qquad (1)$$

Это хорошо известная из школьного курса математики формула для суммы бесконечно убывающей геометрической прогрессии. Дифференцируя (1) почленно, получаем, что производящая функция последовательности натуральных чисел 1, 2, ... , n, ... есть

$$A(x)=1+2x+3x^2+\ldots+nx^{n-1}+\ldots=\frac{1}{(1-x)^2}.$$

С помощью k — кратного дифференцирования (1) получаем

$$\sum_{n=0}^{\infty}\frac{(n+k)!}{n!}x^n=\frac{k!}{(1-x)^{k+1}};$$

$$\sum_{n=0}^{\infty}C_{k+n}^{k}x^n=\frac{1}{(1-x)^{k+1}},\quad k=0,1,2,\ldots. \qquad (2)$$

Полезно также следующее непосредственно проверяемое тождество

$$\frac{1-x^n}{1-x}=1+x+x^2+\ldots+x^{n-1}. \qquad (3)$$

Теперь, вооруженные уже определенными техническими средствами, перейдем к задачам, решение которых методом производящих функций путем прямого перемножения скобок было бы достаточно утомительным.

Какова вероятность при бросании 4 игральных костей выбросить 14 очков?

Эта вероятность равна отношению числа способов выбросить 14 очков к полному числу возможных выпадений 4 различных костей. Подсчет полного числа выпадений 4 различных костей не составляет труда. Так как каждая кость имеет 6 граней, то это число равно 6^4. Сложнее найти число способов, которыми может быть выброшено 14 очков. Ответом на этот вопрос является коэффициент при x^{14} в выражении $(x+x^2+x^3+x^4+x^5+x^6)^4$. Здесь каждая из четырёх скобок, возникающих при расписывании степени в виде произведения, соответствует одной из игральных костей, а стоящие в скобке степени переменной x выражают возможные значения очков на данной кости. После выполнения алгебраических преобразований коэффициенты степенного ряда укажут число способов получения любой суммы от 4 до 24. Поэтому полученный степенной ряд (в данном случае конечный) называется производящей функцией для числа способов выпадения различных сумм в четырёх бросаниях. Используя (3) и (2), найдём эту производящую функцию

$$(x+x^2+x^3+x^4+x^5+x^6)^4 = x^4\cdot(1+x+x^2+x^3+x^4+x^5)^4$$
$$= x^4\cdot(1-x^6)^4\cdot\frac{1}{(1-x)^4} =$$
$$= x^4\cdot\left(\sum_{i=0}^{4} C_4^i(-1)^i x^{6i}\right)\cdot\left(\sum_{n=0}^{\infty} C_{n+3}^3 x^n\right).$$

Легко видеть, что степень x^{14} возникает в двух случаях: когда в первой сумме $i=0$, а во второй $n=10$, и когда в первой сумме $i=1$, а во второй $n=4$. Поэтому коэффициент при x^{14} равен $C_4^0\cdot C_{13}^3 - C_4^1\cdot C_7^3 = 146$. Таким образом, искомая вероятность равна

$$P=\frac{146}{6^4}=\frac{146}{1296}\approx 0{,}11.\blacktriangleleft$$

Чтобы сравнить возможности различных подходов для решения задач перечисления, рассмотрим задачу из раздела 3, решённую там с помощью формулы включения и исключения, и решим её методом производящих функций.

Пусть снова требуется найти число целочисленных решений системы

$$\begin{cases} x_1+x_2+x_3=40, \\ 4\le x_1\le 15, \\ 9\le x_2\le 18, \\ 5\le x_3\le 16. \end{cases}$$

Легко понять, что искомое число решений есть коэффициент при z^{40} после раскрытия скобок в выражении

$$\left(z^4+\ldots+z^{15}\right)\left(z^9+\ldots+z^{18}\right)\left(z^5+\ldots+z^{16}\right).$$

Три перемножаемые скобки соответствуют переменным x_1, x_2, x_3, и выписанное выражение является производящей функцией для числа решений системы

$$\begin{cases} x_1+x_2+x_3=n, \\ 4\le x_1\le 15, \\ 9\le x_2\le 18, \\ 5\le x_3\le 16, \end{cases}$$

так как при любом целом $n\ge 0$ коэффициент при z^n равен числу решений. Найдём коэффициент при z^{40}.

$$\left(z^4+\ldots+z^{15}\right)\left(z^9+\ldots+z^{18}\right)\left(z^5+\ldots+z^{16}\right)=$$

$$=z^4\left(1+\ldots+z^{11}\right)z^9\left(1+\ldots+z^9\right)z^5\left(1+\ldots+z^{11}\right)=$$

$$=z^{18}\cdot\frac{1-z^{12}}{1-z}\cdot\frac{1-z^{10}}{1-z}\cdot\frac{1-z^{12}}{1-z}=$$

$$=z^{18}\cdot\frac{1}{(1-z)^3}\cdot\left(1-z^{10}-2z^{12}+2z^{22}+z^{24}-z^{34}\right)=$$

$$=z^{18}\cdot\left(\sum_{n=0}^{\infty}C_{n+2}^2\,z^n\right)\cdot(1-z^{10}-2z^{12}+2z^{22}+z^{24}-z^{34}).$$

В коэффициент при z^{40} дают вклад значения $n=22,12,10,0$. Поэтому число решений равно

$$C_{24}^2-C_{14}^2-2C_{12}^2+2C_2^2=\frac{24\cdot23}{2}-\frac{14\cdot13}{2}-2\frac{12\cdot11}{2}+2=55.$$

Разбиения чисел. Обратимся к уже рассматривавшимся в разделе 1 разбиениям чисел на суммы, используя теперь в качестве метода исследования производящие функции, впервые использованные в этой задаче Эйлером [21]. Пусть $P\left(n,\{a,b,c,\ldots\}\right)$ число разбиений натурального числа n на суммы, слагаемые в которых берутся из множества $\{a,b,c,\ldots\}$ с возможными повторениями. Производящая функция для $P\left(n,\{a,b,c,\ldots\}\right)$ есть

$$\sum_{n=0}^{\infty}P\left(n,\{a,b,c,\ldots\}\right)x^n=$$

$$=(1+x^a+x^{2a}+x^{3a}+x^{4a}+\ldots)(1+x^b+x^{2b}+\ldots)(1+x^c+x^{2c}+\ldots)\ldots=$$

$$=(1-x^a)^{-1}(1-x^b)^{-1}(1-x^c)^{-1}\ldots.$$

Производящая функция для числа разбиений на натуральные слагаемые есть

$$\sum_{n=0}^{\infty}P(n)x^n=(1-x)^{-1}\left(1-x^2\right)^{-1}\left(1-x^3\right)^{-1}\ldots.$$

Производящая функция для числа разбиений на нечётные слагаемые есть

$$\sum_{n=0}^{\infty}P\left(n,\{1,3,5,\ldots\}\right)x^n=(1-x)^{-1}\left(1-x^3\right)^{-1}\left(1-x^5\right)^{-1}\ldots.$$

Пусть теперь $P'(n,\{a,b,c,\ldots\})$ число разбиений натурального числа n на суммы, слагаемые в которых берутся из множества $\{a,b,c,\ldots\}$ без повторений. Для производящей функции имеем

$$\sum_{n=0}^{\infty} P'(n,\{a,b,c,\ldots\})x^n = (1+x^a)(1+x^b)(1+x^c)\ldots .$$

Производящая функция для числа разбиений на натуральные слагаемые без повторений

$$\sum_{n=0}^{\infty} P'(n)x^n = (1+x)(1+x^2)(1+x^3)\ldots .$$

Производящая функция для числа разбиений на нечётные слагаемые без повторений

$$\sum_{n=0}^{\infty} P'(n,\{1,3,5,\ldots\})x^n = (1+x)(1+x^3)(1+x^5)\ldots .$$

С помощью производящих функций можно получить следующий интересный результат о числе разбиений.

Для каждого натурального n число его разбиений на слагаемые без повторений равно числу его разбиений на нечётные слагаемые с возможными повторениями: $P'(n) = P(n,\{1,3,5,\ldots\})$.

$$\sum_{n=0}^{\infty} P'(n)x^n = (1+x)(1+x^2)(1+x^3)\ldots =$$

$$= \frac{(1-x)(1+x)(1-x^2)(1+x^2)(1-x^3)(1+x^3)\ldots}{(1-x)(1-x^2)(1-x^3)\ldots} =$$

$$= \frac{(1-x^2)(1-x^4)(1-x^6)\ldots}{(1-x)(1-x^2)(1-x^3)(1-x^4)\ldots} = \frac{1}{(1-x)(1-x^3)(1-x^5)} =$$

$$= \sum_{n=0}^{\infty} P(n,\{1,3,5,\ldots\})x^n$$

В качестве ещё одного любопытного приложения производящих функций докажем единственность представления натурального числа в двоичной системе счисления.

$$\sum_{n=0}^{\infty} P'(n,\{1,2,4,8,\ldots\})x^n = (1+x)(1+x^2)(1+x^4)\ldots =$$

$$= \frac{1-x^2}{1-x}\frac{1-x^4}{1-x^2}\frac{1-x^8}{1-x^4}\ldots = \frac{1}{1-x} = \sum_{n=0}^{\infty} x^n .$$

Формула Добинского. Информация о комбинаторных числах в методе производящих функций извлекается путём аналитической работы с бесконечными рядами, порождаемыми последовательностями этих чисел, что требует определённых навыков такой работы и аналитической техники. Продемонстрируем мощь метода производящих функций на сравнительно несложном в техническом отношении, но классическом примере вывода формулы Добинского (1.7) для чисел Белла, которая в разделе 1 была приведена без доказательства.

Заметим, прежде всего, что

$$x^k e^x = x^k \sum_{i=0}^{\infty} \frac{x^i}{i!} = \sum_{i=0}^{\infty} \frac{x^{i+k}}{i!} = \sum_{i=0}^{\infty} \frac{(i+k)!}{(i+k)!} \frac{x^{i+k}}{i!} =$$

$$= \sum_{i=0}^{\infty} \frac{(i+k)!}{i!} \frac{x^{i+k}}{(i+k)!} = \sum_{m=k}^{\infty} A_m^k \frac{x^m}{m!} = \sum_{m=0}^{\infty} A_m^k \frac{x^m}{m!},$$

так как $A_m^k = 0$ при $m \le k-1$. Рассматривая теперь для чисел Стирлинга производящую функцию $\sum_{k=0}^{\infty} S_n^k x^k$ и используя доказанное соотношение, а также соотношение (1.12), получаем

$$e^x \sum_{k=0}^{\infty} S_n^k x^k = \sum_{k=0}^{\infty} S_n^k x^k e^x = \sum_{k=0}^{\infty} S_n^k \sum_{m=0}^{\infty} A_m^k \frac{x^m}{m!} = \sum_{m=0}^{\infty} \frac{x^m}{m!} \sum_{k=0}^{\infty} A_m^k S_n^k = \sum_{m=0}^{\infty} \frac{x^m}{m!} m^n .$$

Подставляя в полученное тождество $x = 1$, получаем формулу Добинского

$$B_n = \frac{1}{e} \sum_{m=0}^{\infty} \frac{m^n}{m!} .$$

Рекуррентные последовательности. Производящие функции могут быть эффективно использованы для исследования последовательностей, заданных *линейными рекуррентными соотношениями*. Первые k членов такой последовательности задаются, а каждый последующий член находится по k членам, ему предшествующим, с помощью заданного линейного соотношения. Термин «рекуррентный» введён Муавром, впервые применившим метод производящих функций для решения линейных рекуррентных соотношений. Абрахам де Муавр (1667–1754), француз по национальности, в результате религиозного преследования гугенотов в возрасте двадцати одного года был вынужден покинуть Францию и поселиться в Лондоне. Помимо метода производящих функций им был сделан целый ряд других важных математических открытий, в частности, получена знаменитая предельная теорема Муавра—Лапласа в теории вероятностей.

Простейшими линейными рекуррентными последовательностями являются арифметическая прогрессия, задаваемая соотношением $a_{n+1} = a_n + d$, и

геометрическая прогрессии, задаваемая соотношением $b_{n+1} = q\,b_n$. У этих рекуррентных последовательностей каждый последующий член определяется предыдущим, а первый член задаётся, поэтому получение формулы для n-го члена не составляет труда: $a_n = a_1 + (n-1)d$, $b_n = b_1\,q^{n-1}$. Задача становится нетривиальной, если каждый последующий член определяется двумя или большим числом предыдущих членов.

Рассмотрим пример рекуррентной последовательности $\{u_n\} = u_0, u_1, u_2, \ldots$, у которой каждый член, начиная с третьего, выражается через два предыдущих с помощью соотношения $u_{n+2} = 4u_{n+1} - 3u_n$, а два первых члена равны $u_0 = 8$, $u_1 = 10$. С помощью рекуррентного соотношения, используя заданные значения u_0 и u_1, получаем, что $u_2 = 16$, $u_3 = 34$ и т. д. Задача состоит в том, чтобы найти в явном виде формулу n-го члена.

Помножив обе части рекуррентного соотношения на x^{n+2} и просуммировав по n, получим

$$\sum_{n=0}^{\infty} u_{n+2}x^{n+2} = 4x\sum_{n=0}^{\infty} u_{n+1}x^{n+1} - 3x^2\sum_{n=0}^{\infty} u_n x^n .$$

Отсюда для производящей функции

$$U(x) = u_0 + u_1 x + u_2 x^2 + \ldots$$

имеем соотношение:

$$U(x) - u_0 - u_1 x = 4x(U(x) - u_0) - 3x^2 U(x);$$

$$U(x) - 8 - 10x = 4x(U(x) - 8) - 3x^2 U(x).$$

Разрешая его относительно $U(x)$, получаем

$$U(x) = \frac{-22x+8}{3x^2 - 4x + 1} = \frac{-22x+8}{(x-1)(3x-1)}.$$

Разложим полученное дробно-рациональное выражение на простейшие дроби

$$\frac{-22x+8}{(x-1)(3x-1)} = \frac{a}{x-1} + \frac{b}{3x-1},$$

где a и b — подлежащие определению числа. Так как

$$\frac{a}{x-1} + \frac{b}{3x-1} = \frac{(3a+b)x - (a+b)}{(x-1)(3x-1)},$$

находим неизвестные a и b из системы уравнений

$$\begin{cases} 3a + b = -22 \\ a + b = -8, \end{cases}$$

что даёт $a=-7$, $b=-1$, и разложение приобретает вид

$$U(x)=\frac{-22x+8}{(x-1)(3x-1)}=\frac{-7}{x-1}+\frac{-1}{3x-1}=\frac{7}{1-x}+\frac{1}{1-3x}.$$

Используя (1), имеем

$$\frac{7}{1-x}=7(1+x+x^2+\ldots)$$

и

$$\frac{1}{1-3x}=1+(3x)+(3x)^2+\ldots.$$

Таким образом, окончательно получаем разложение $U(x)$ в степенной ряд в виде

$$U(x)=7(1+x+x^2+\ldots)+1+(3x)+(3x)^2+\ldots=\sum_{n=0}^{\infty}(7+3^n)x^n.$$

Это позволяет выписать формулу n-го члена:

$$u_n=7+3^n.$$

Использованный здесь метод является общим приёмом для решения линейных рекуррентных соотношений, суть которого можно выразить следующим образом:

1) *из заданного рекуррентного соотношения получаем линейное уравнение для производящей функции, решая которое находим производящую функцию последовательности;*
2) *раскладывая производящую функцию в ряд, находим члены искомой последовательности как коэффициенты этого степенного ряда.*

Рассмотрим теперь, как рекуррентные соотношения возникают в реальных перечислительных задачах. Пусть требуется найти число F_n двоичных последовательностей длины n, не содержащих двух единиц подряд. При $n=1$ имеются две таких последовательности: (0) и (1). При $n=2$ таких последовательностей три: (0,0), (0,1), (1,0). Поэтому $F_1=2$, $F_2=3$.

Разобьём искомое множество последовательностей длины n на 2 подмножества: последовательности, начинающиеся с 0, и последовательности, начинающиеся с 1. Последовательности первого типа не имеют каких-либо дополнительных ограничений на последующие $n-1$ символов. Поэтому их F_{n-1}. Последовательности второго типа на второй позиции обязаны содержать 0, а на последующие $n-2$ символов нет каких-либо ограничений, поэтому их F_{n-2}. Это приводит к рекуррентному соотношению

$$F_n=F_{n-1}+F_{n-2}, \qquad (4)$$

т. е. каждый член последовательности, начиная с третьего, равен сумме двух предыдущих членов. Положим $F_0 = 1$. Тогда рекуррентное соотношение будет выполняться, начиная со второго члена. Вместе с начальными данными $F_0 = 1$ и $F_1 = 2$ оно позволяет найти любой член последовательности $F_0, F_1, F_2, \ldots, F_n, \ldots$. Вот первые семь членов: 1, 2, 3, 5, 8, 13, 21, ...

Данная последовательность называется последовательностью Фибоначчи, по имени впервые рассмотревшего её итальянского математика XIII века. Она возникла в придуманной им своеобразной задаче о размножении кроликов. Пара кроликов, появившись на свет, даёт в качестве приплода каждый месяц, исключая первый месяц после рождения, ещё пару кроликов. Некто покупает одну взрослую пару и ежемесячно считает число имеющихся у него пар, которое может быть найдено с помощью рекуррентного соотношения (4).

Для получения формулы n-го члена в явном виде домножим обе части рекуррентного соотношения (4) на x^n и просуммируем по n от 2 до ∞:

$$\sum_{n=2}^{\infty} F_n x^n = \sum_{n=2}^{\infty} F_{n-1} x^n + \sum_{n=2}^{\infty} F_{n-2} x^n .$$

Это позволяет для производящей функции

$$F(x) = F_0 + F_1 x + F_2 x^2 + \ldots + F_n x^n + \ldots$$

получить соотношение

$$F(x) - 1 - 2x = x\left(F(x) - 1\right) + x^2 F(x).$$

Разрешая это соотношение относительно $F(x)$, получаем

$$F(x) = -\frac{x+1}{x^2 + x - 1}.$$

Разлагаем данное рациональное выражение на простейшие дроби:

$$x^2 + x - 1 = (x - x_1)(x - x_2), \text{ где } x_1 = \frac{-1+\sqrt{5}}{2},\ x_2 = \frac{-1-\sqrt{5}}{2};$$

$$F(x) = \frac{a}{x - x_1} + \frac{b}{x - x_2} = -\frac{x-1}{(x - x_1)(x - x_2)};$$

$$a(x - x_2) + b(x - x_1) = -x - 1.$$

Подставляя $x = x_1$, получаем

$$a = -\frac{x_1 + 1}{x_1 - x_2} = -\frac{1+\sqrt{5}}{2\sqrt{5}}.$$

Подставляя $x = x_2$, получаем

$$b = -\frac{x_2 + 1}{x_1 - x_2} = -\frac{1-\sqrt{5}}{2\sqrt{5}}.$$

Для производящей функции $F(x)$ имеем

$$F(x) = -\frac{1}{x_1}\cdot\frac{a}{1-\frac{x}{x_1}} - \frac{1}{x_2}\cdot\frac{b}{1-\frac{x}{x_2}} = -\frac{a}{x_1}\sum_{n=0}^{\infty}\left(\frac{x}{x_1}\right)^n - \frac{b}{x_2}\sum_{n=0}^{\infty}\left(\frac{x}{x_2}\right)^n =$$

$$= -\sum_{n=0}^{\infty}\left(\frac{a}{x_1^{n+1}} + \frac{b}{x_2^{n+1}}\right)x^n.$$

Отсюда n-ый член последовательности Фибоначчи равен

$$F_n = -\frac{a}{x_1^{n+1}} - \frac{b}{x_2^{n+1}} = \frac{1}{\sqrt{5}}\left(\left(\frac{\sqrt{5}+1}{2}\right)^{n+2} + (-1)^{n+1}\left(\frac{\sqrt{5}-1}{2}\right)^{n+2}\right).$$

Заметим, что второй член в квадратных скобках стремится к нулю с ростом n. Поэтому для нахождения n-го члена последовательности достаточно найти вклад, даваемый первым членом, и округлить результат до ближайшего целого числа.

Несмотря на простоту исходного рекуррентного соотношения, формула общего члена оказалась не слишком простой. Во времена Фибоначчи в математике еще не было средств для её нахождения. Она была найдена значительно позже. Сам же Леонардо из Пизы наслаждался последовательным вычислением числа пар кроликов с помощью своего рекуррентного соотношения. К его восторгу следует отнестись с уважением, ведь это была первая в истории математики рекуррентная последовательность после известных с античных времён арифметической и геометрической прогрессий. Последовательность Фибоначчи обладает многими интересными свойствами, которые интенсивно изучались математиками последующих веков.

Производящая функция частичных сумм. Пусть $a_0, a_1, \ldots, a_n, \ldots$ — последовательность, $A(x) = \sum_{n=0}^{\infty} a_n x^n$ — её производящая функция. Рассмотрим последовательность частичных сумм $s_0, s_1, \ldots, s_n, \ldots$, где $s_n = \sum_{i=0}^{n} a_i$, с производящей функцией $S(x) = \sum_{n=0}^{\infty} s_n x^n$, и найдём связь между производящими функциями $A(x)$ и $S(x)$. Для последовательности $\{s_i\}$ имеет место очевидное рекуррентное соотношение: $s_{n+1} = s_n + a_{n+1}$. Домножив обе его части на x^{n+1} и просуммировав по n от 0 до ∞, получаем

$$\sum_{n=0}^{\infty} s_{n+1}x^{n+1} = x\sum_{n=0}^{\infty} s_n x^n + \sum_{n=0}^{\infty} a_{n+1}x^{n+1};$$

$$S(x) - a_0 = x \cdot S(x) + A(x) - a_0.$$

Отсюда находим связь между производящими функциями:

$$S(x) = \frac{A(x)}{1-x}. \qquad (5)$$

С помощью полученного соотношения найдём формулу для суммы квадратов n первых натуральных чисел. Хорошо известно, что

$$1^2 + 2^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

Доказательство этой формулы является классическим примером применения метода математической индукции. Однако сама формула при этом должна быть каким-то образом угадана. Метод производящих функций позволяет легко получить эту формулу и не требует никаких угадываний.

Пусть $A(x)$ — производящая функция последовательности квадратов натуральных чисел:

$$A(x) = \sum_{i=0}^{\infty} i^2 x^i = \sum_{i=0}^{\infty} (i(i-1)+i)x^i = \sum_{i=0}^{\infty} i(i-1)x^i + \sum_{i=0}^{\infty} ix^i =$$

$$= x^2 \sum_{i=0}^{\infty} i(i-1)x^{i-2} + x\sum_{i=0}^{\infty} ix^{i-1} = x^2\left(\sum_{i=0}^{\infty} x^i\right)'' + x\left(\sum_{i=0}^{\infty} x^i\right)' =$$

$$= x^2\left(\frac{1}{1-x}\right)'' + x\left(\frac{1}{1-x}\right)' = \frac{2x^2}{(1-x)^3} + \frac{x}{(1-x)^2} = \frac{x^2+x}{(1-x)^3}.$$

Теперь, используя (5) и (2), находим производящую функцию суммы квадратов

$$S(x) = \frac{A(x)}{1-x} = \frac{x^2+x}{(1-x)^4} = (x^2+x)\sum_{n=0}^{\infty} C_{n+3}^3 x^n =$$

$$= x^2 \sum_{n=0}^{\infty} C_{n+3}^3 x^n + x\sum_{n=0}^{\infty} C_{n+3}^3 x^n = \sum_{n=0}^{\infty} C_{n+3}^3 x^{n+2} + \sum_{n=0}^{\infty} C_{n+3}^3 x^{n+1} =$$

$$= \sum_{n=2}^{\infty} C_{n+1}^3 x^n + \sum_{n=1}^{\infty} C_{n+2}^3 x^n = \sum_{n=0}^{\infty} C_{n+1}^3 x^n + \sum_{n=0}^{\infty} C_{n+2}^3 x^n =$$

$$= \sum_{n=0}^{\infty} \left(C_{n+1}^3 + C_{n+2}^3\right) x^n.$$

$$\sum_{i=0}^{n} i^2 = C_{n+1}^3 + C_{n+2}^3 = \frac{(n+1)!}{(n-2)!3!} + \frac{(n+2)!}{(n-1)!3!} =$$

$$= \frac{(n-1)n(n+1) + n(n+1)(n+2)}{6} = \frac{n(n+1)(2n+1)}{6}.$$

Производящая функция свёртки. Пусть заданы две последовательности $a_0, a_1, a_2, \ldots, a_n, \ldots$ и $b_0, b_1, b_2, \ldots, b_n, \ldots$. Последовательность a_0b_0, $(a_0b_1 + a_1b_0)$, $(a_0b_2 + a_1b_1 + a_2b_0)$, ..., $\sum_{i=0}^{n} a_i b_{n-i}$, ... называется их *свёрткой*. Перемножая производящие функции $A(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n + \ldots$ и $B(x) = b_0 + b_1x + b_2x^2 + \ldots + b_nx^n + \ldots$, получаем

$$A(x) \cdot B(x) = a_0b_0 + (a_0b_1 + a_1b_0)x + (a_0b_2 + a_1b_1 + a_2b_0)x^2 + \ldots + \left(\sum_{i=0}^{n} a_i b_{n-i}\right)x^n + \ldots,$$

т. е. **производящая функция свёртки последовательностей является произведением их производящих функций.**

Воспользовавшись полученным соотношением, решим задачу о разбиении выпуклого многоугольника на треугольники диагоналями, не пересекающимися внутри многоугольника. Такое разбиение называется триангуляцией многоугольника. Сколькими способами возможно триангулировать выпуклый многоугольник? Эта задача, как и ряд других задач, с которыми читатель ещё встретится на страницах этой книги, принадлежит Эйлеру[1].

Обозначив число триангуляций $(n+2)$-угольника через C_n, имеем: $C_1 = 1$, $C_2 = 2$, $C_3 = 5$, $C_4 = 14$, ... На рис. 1 представлены все 5 возможных триангуляций пятиугольника:

Найдём рекуррентное соотношение для C_n. Пусть при некоторой триангуляции в один треугольник с вершинами 1 и $(n+2)$ попадает вершина i. Найдем число таких триангуляций. Если $i = 2$ или $i = n+1$, то число три-

---

[1] Леонард Эйлер (1707–1783) был крупнейшим математиком и механиком XVIII века. Он внёс существенный вклад во все разделы современной ему математики и заложил ряд новых, одинаково успешно занимаясь как теоретическими, так и прикладными вопросами. Эйлер придал математическому анализу ту форму, в которой он и поныне преподаётся в технических вузах. Он также первым проявил интерес к задачам, которые сегодня принято относить к области дискретной математики.

Эйлер родился в Швейцарии, но на протяжении многих лет жил и работал в России, являясь Российским академиком. В конце жизни Эйлер полностью ослеп, но это не снизило его научной продуктивности. Когда Эйлер умер, непременный секретарь Парижской академии наук Кондорсе сказал: «Эйлер перестал жить и вычислять».

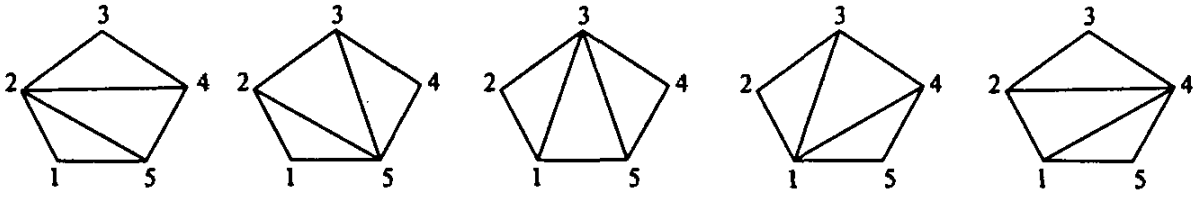


Рис. 1

ангуляций равно C_{n-1}. Для $3 \le i \le n$ число триангуляций равно произведению числа триангуляций i-угольника на число триангуляций $(n-i+3)$-угольника, т. е. равно $C_{i-2}C_{n-i+1}$. Таким образом, полагая $C_0 = 1$, получаем для C_n рекуррентное соотношение:

$$C_n = \sum_{i=2}^{n+1} C_{i-2} C_{n-i+1} = \sum_{i=0}^{n-1} C_i C_{n-i-1}. \qquad (6)$$

В правой части (6) стоит $(n-1)$-й член свёртки последовательности C_n с самой собою. Домножив обе части (6) на x^n и просуммировав по n от 1 до ∞, получим

$$\sum_{n=1}^{\infty} C_n x^n = x \sum_{n=1}^{\infty} \left(\sum_{i=0}^{n-1} C_i C_{n-i-1} \right) x^{n-1}.$$

Это даёт для производящей функции $C(x) = C_0 + C_1 x + C_2 x^2 + \ldots$ соотношение

$$C(x) - C_0 = xC^2(x);$$

$$xC^2(x) - C(x) + 1 = 0.$$

Решением квадратного уравнения, удовлетворяющим условию $C(0) = 1$, является

$$C(x) = \frac{1 - \sqrt{1-4x}}{2x}.$$

Нашей дальнейшеё целью является разложение этого выражения в ряд. Имеем

$$\sqrt{1-4x} = 1 - 2x - \sum_{n=2}^{\infty} \frac{1 \cdot 3 \cdot \ldots \cdot (2n-3) \cdot 4^n}{2^n n!} x^n = 1 - 2x - \sum_{n=2}^{\infty} \frac{1 \cdot 3 \cdot \ldots \cdot (2n-3) \cdot 2^n}{n!} x^n =$$

$$= 1 - 2x - \sum_{n=2}^{\infty} \frac{1 \cdot 3 \cdot \ldots \cdot (2n-3) \cdot 2^n (n-1)!}{n!\,(n-1)!} x^n =$$

$$= 1 - 2x - \sum_{n=2}^{\infty} \frac{2}{n} \cdot \frac{1 \cdot 3 \cdot \ldots \cdot (2n-3) \cdot 2^{n-1} (n-1)!}{(n-1)!\,(n-1)!} x^n =$$

$$=1-2x-\sum_{n=2}^{\infty}\frac{2}{n}\cdot\frac{1\cdot3\cdot\ldots\cdot(2n-3)\cdot2\cdot4\cdot\ldots\cdot(2n-2)}{(n-1)!\,(n-1)!}x^n=$$

$$=1-2x-\sum_{n=2}^{\infty}\frac{2}{n}\cdot\frac{(2n-2)!}{(n-1)!\,(n-1)!}x^n=1-2x-\sum_{n=2}^{\infty}\frac{2}{n}C_{2n-2}^{n-1}x^n=1-2\sum_{n=1}^{\infty}\frac{1}{n}C_{2n-2}^{n-1}x^n.$$

Следовательно

$$C(x)=\sum_{n=1}^{\infty}\frac{1}{n}C_{2n-2}^{n-1}x^{n-1}=\sum_{n=0}^{\infty}\frac{1}{n+1}C_{2n}^{n}x^n.$$

Отсюда получаем

$$C_n=\frac{1}{n+1}C_{2n}^{n}. \qquad (7)$$

Числа C_n возникают во многих задачах дискретной математики и имеют специальное название. Их называют числами Каталана по имени бельгийского математика XIX века.

В качестве другого примера появления чисел Каталана (7) рассмотрим множество векторов длины $2n$, компонентами которых являются $+1$ и -1, причём число единиц со знаком плюс равно числу единиц со знаком минус. Таких (±1)-векторов, очевидно, C_{2n}^{n}. Поставим теперь задачу найти среди них число векторов, у которых сумма любого числа первых компонент неотрицательна. Будем обозначать множество таких векторов через T_n, а их число $|T_n|=\tau_n$.

При $n=1$ имеем единственный такой вектор: $T_1=\{(+1,-1)\}$. При $n=2$ таких векторов два: $T_2=\{(+1,+1,-1,-1),\ (+1,-1,+1,-1)\}$. При $n=3$ подобных векторов уже пять:

$$T_3=\{(+1,+1,+1,-1,-1,-1),\ (+1,+1,-1,+1,-1,-1),\ (+1,+1,-1,-1,+1,-1),$$
$$(+1,-1,+1,+1,-1,-1),\ (+1,-1,+1,-1,+1,-1)\}.$$

Имеем $\tau_1=1$, $\tau_2=2$, $\tau_3=5$. Это есть начальные члены последовательности Каталана $C_1, C_2, C_3, \ldots$. Покажем, что для любого n справедливо $\tau_n=C_n$.

Обозначим через $T_n'\subseteq T_n$ подмножество тех векторов из T_n, у которых для любого i, $1\le i\le n-1$, сумма первых $2i$ координат строго больше нуля, $|T_n'|=\tau_n'$.

Число векторов из T_n, у которых для некоторого i, $1\le i\le n-1$, сумма первых $2i$ координат вектора первый раз обращается в нуль, очевидно, равно $\tau_i'\cdot\tau_{n-i}$. Поэтому, положив $\tau_0=1$, имеем

$$\tau_n = \sum_{i=1}^{n-1} \tau'_i \tau_{n-i} + \tau'_n = \sum_{i=1}^{n} \tau'_i \tau_{n-i} \,. \qquad (8)$$

Первая координата любого вектора из T'_n всегда равна $+1$, а последняя — (-1). Если эти координаты отбросить, то, как легко видеть, получится вектор из T_{n-1}. Обратно, приписав к любому вектору из T_{n-1} в качестве первой координаты $+1$, а в качестве последней — (-1), получим вектор из T'_n. Отсюда $\tau'_n = \tau_{n-1}$. Это позволяет переписать (8) в виде

$$\tau_n = \sum_{i=1}^{n} \tau_{i-1} \tau_{n-i} = \sum_{i=0}^{n-1} \tau_i \tau_{n-i-1} \,. \qquad (9)$$

Но рекуррентное соотношение (9) совпадает с рекуррентным соотношением (6), определяющим числа Каталана. Отсюда следует, что

$$\tau_n = C_n = \frac{1}{n+1} C_{2n}^n \,.$$

Решение последней задачи показывает широту приложений метода производящих функций, но его вряд ли можно назвать элементарным. К счастью, для этой задачи существует и неожиданно простое геометрическое решение.

Метод траекторий. Будем рассматривать $(2n)$-мерный (± 1)- вектор как одномерное блуждание точки вдоль оси y. Точка, в начальный момент находящаяся в нуле, в дискретные моменты времени $t = 1, 2, \ldots, 2n$ перемещается на единицу в положительном или отрицательном направлении, в зависимости от знака соответствующей компоненты (± 1)-вектора. Требуется найти число таких блужданий, при которых в момент времени $t = 2n$ точка возвратится в нуль, ни разу не побывав при этом в отрицательной области. Сделав «развёртку» такого одномерного движения с помощью оси времени t, получим его траекторию. На рис. 2 представлена такая траектория, соответствующая вектору $(+1, -1, +1, +1, -1, +1, -1, -1)$.

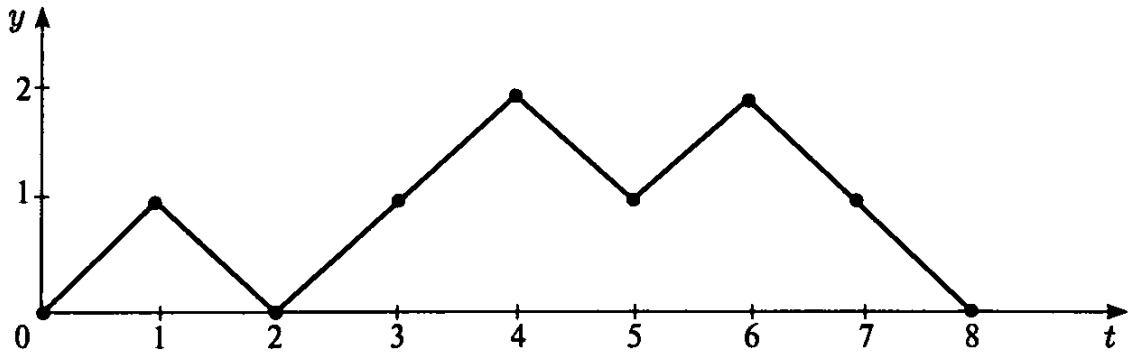


Рис. 2

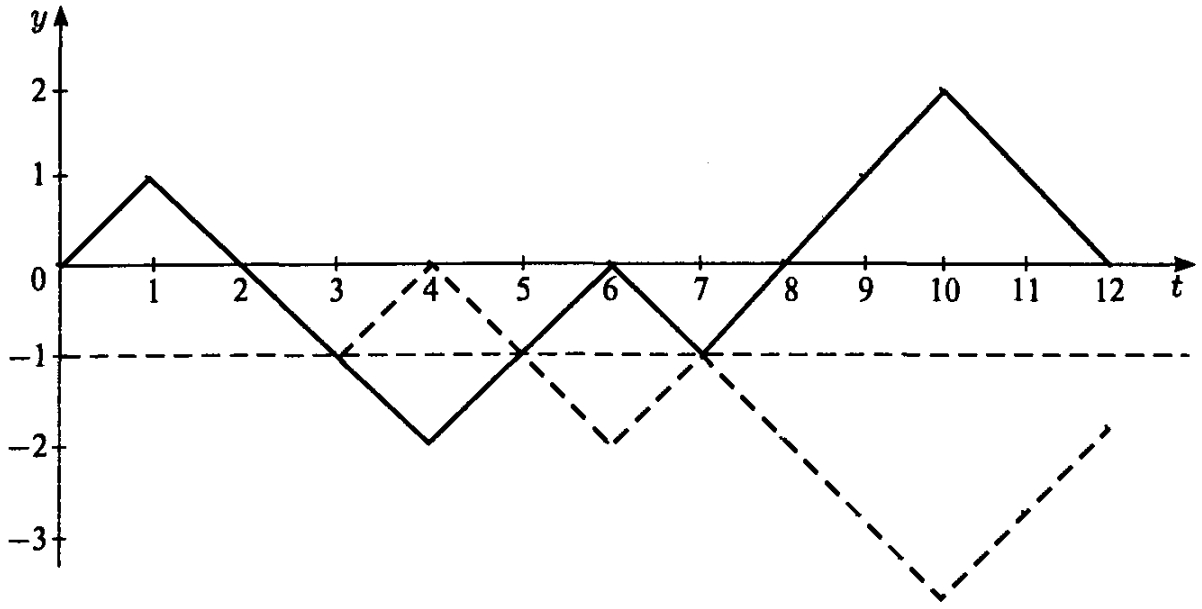


Рис. 3

Таким образом, нужно найти число путей, ведущих из точки $(0,0)$ в точку $(2n,0)$ и не заходящих в отрицательную область. Так как полное число путей из точки $(0,0)$ в точку $(2n,0)$ равно C_{2n}^{n}, то для решения задачи достаточно из C_{2n}^{n} вычесть число путей, заходящих в отрицательную область, которое определяется совсем просто.

В самом деле, пусть имеется путь из $(0,0)$ в $(2n,0)$, заходящий в отрицательную область. Отразив симмстрично относительно прямой $y=-1$ его часть после первого попадания на эту прямую, получим путь из точки $(0,0)$ в точку $(2n,-2)$. На рис. 3 показано отображение заходящего в отрицательную область пути из $(0,0)$ в $(12,0)$ в путь из $(0,0)$ в $(12,-2)$. Сплошной линией показан исходный путь, а прерывистой — его отражение.

Как легко понять из рисунка, в общем случае данным отражением устанавливается взаимно однозначное соответствие между множеством всех путей из $(0,0)$ в $(2n,-2)$ и множеством путей из $(0,0)$ в $(2n,0)$, заходящих в отрицательную область. В самом деле, любой путь из $(0,0)$ в $(2n,-2)$ хотя бы раз пересекает прямую $y=-1$, поэтому, отразив его часть после первого пересечения этой прямой симметрично относительно неё, получим путь из $(0,0)$ в $(2n,0)$, заходящий в отрицательную область. Так как имеется C_{2n}^{n+1} путей из $(0,0)$ в $(2n,-2)$, число не заходящих в отрицательную область путей из $(0,0)$ в $(2n,0)$ равно

$$C_{2n}^{n}-C_{2n}^{n+1}=\frac{1}{n+1}C_{2n}^{n}=C_n.$$

Вопросы для самопроверки

1. Найти производящую функцию для числа способов выбрасывания n очков при одновременном бросании k игральных костей.

   а) $\frac{1}{(1-x)^k}$; б) $(x+x^2+\ldots+x^k)^6$; в) $(x+x^2+\ldots+x^6)^k$.

2. Найти производящую функцию последовательности C_n^2, $n=0,1,2,\ldots$.

   а) $\frac{x^2}{(1-x)^3}$; б) $\frac{1}{(1-x)^3}$; в) $\frac{1}{(1-x)^4}$.

3. Сколькими способами можно расставить в очередь $2n$ человек, n мужчин и n женщин так, в любом начальном отрезке очереди число мужчин не превышало числа женщин?

   а) 2^n; б) C_{2n}^n; в) $\frac{1}{n+1}C_{2n}^n$.

Ответы: 1 — в, 2 — а, 3 — в.

1.6. Перечисление классов эквивалентности. Теория Пойа

Рассмотрим задачу о числе раскрасок граней куба в 3 цвета: белый (white), синий (blue) и красный (red). Так как есть три возможности для раскраски каждой из 6 граней, то всего раскрасок $3^6 = 729$. Далее, можно найти число раскрасок, использующих i раз белый, j раз синий и k раз красный цвет ($i+j+k=6$). Для этого достаточно раскрыть $(w+b+r)^6$ с помощью полиномиальной формулы

$$(w+b+r)^6 = \sum_{\substack{i,j,k\geq 0 \\ i+j+k=6}} \frac{6!}{i!\,j!\,k!} w^i b^j r^k .$$

При этом каждой грани куба ставится в соответствие скобка $(\omega+b+r)$, и выбор цвета при раскраске грани соответствует выбору буквы при раскрытии скобок. Таким образом, имеется $\frac{6!}{i!\,j!\,k!}$ раскрасок, использующих i раз белый, j раз синий и k раз красный цвет.

Эти формулы, однако, справедливы лишь в том случае, если все грани куба различны, другими словами, куб занимает определённое положение в пространстве. Если же куб можно поворачивать, совмещая его с самим собой, то различные раскраски могут переходить друг в друга, совмещаясь. При этом естественно считать совместимые раскраски одинаковыми и ставить задачу нахождения числа различных раскрасок. На математическом языке это можно выразить следующим образом.

Группа самосовмещающих вращений куба действует как группа перестановок на множестве из 3^6 раскрашенных кубов. Совместимость различных раскрасок при поворотах куба является отношением эквивалентности. Классы эквивалентности принято называть в данном случае орбитами. Таким образом, требуется найти полное число орбит, а также число орбит, использующих заданное число раз каждый из цветов. Для решения этой задачи воспользуемся классическим результатом о числе орбит, возникающих при действии группы на конечном множестве, который носит имя Уильяма Бернсайда (1852–1927), английского математика, внёсшего значительный вклад в развитие теории групп конечного порядка.

Пусть группа G действует как группа подстановок на конечном множестве X, элементы которого будем называть точками. Результат действия подстановки $g \in G$ на точку $x \in X$ будем обозначать через $(x)g$, т. е. писать точку в скобках слева от действующего на неё элемента группы с тем, чтобы сохранить порядок последовательных действий элементов группы слева направо. Точку $x \in X$ назовём стационарной для элемента $g \in G$, если $(x)g = x$. Пусть X_g — множество стационарных для g точек, $|X_g|$ и $|G|$ — соответственно мощность множества X_g и порядок группы G.

Лемма Бернсайда. Число орбит N при действии группы G на конечном множестве точек X равно среднему по группе числу стационарных точек множества X

$$N = \frac{1}{|G|} \sum_{g \in G} |X_g|.$$

Прежде чем давать доказательство, проиллюстрируем лемму примерами. Пусть множество X состоит из 5 точек: $X = \{1,2,3,4,5\}$, а действующая на этом множестве группа G является циклической группой, порождённой подстановкой шестого порядка $g = (1,2,3)(4,5)$. Ясно, что образующие циклы множества $\{1,2,3\}$ и $\{4,5\}$ и будут в данном случае орбитами действия группы.

Выпишем все 6 подстановок данной циклической группы G:

$$\begin{aligned} e = \; & g^0 = (1)(2)(3)(4)(5); \\ & g^1 = (1,2,3)(4,5); \\ & g^2 = (1,3,2)(4)(5); \\ & g^3 = (1)(2)(3)(4,5); \\ & g^4 = (1,2,3)(4)(5); \\ & g^5 = (1,3,2)(4,5). \end{aligned}$$

Просуммировав число стационарных точек по всем перестановкам и разделив полученную сумму на порядок группы, получаем (5 + 0 + 2 + 3 + + 2 + 0) / 6 = 2, в полном соответствии с леммой.

Рассмотрим теперь пример, более близкий к нашей задаче о раскраске граней куба. Пусть вершины квадрата раскрашиваются двумя цветами: красным (red) и синим (blue). Всего таких раскрасок $2^4 = 16$. Квадрат может поворачиваться относительно центра на углы 0°, 90°, 180°, 270°, самосовмещаясь. При этом одна раскраска может переходить в другую. Требуется найти число различных раскрасок с учётом самосовмещений, т. е. число орбит действия группы поворотов на множестве 16 раскрашенных квадратов. Вот эти 16 раскрасок, обозначенных как $x_1 \div x_{16}$ (рис. 1):

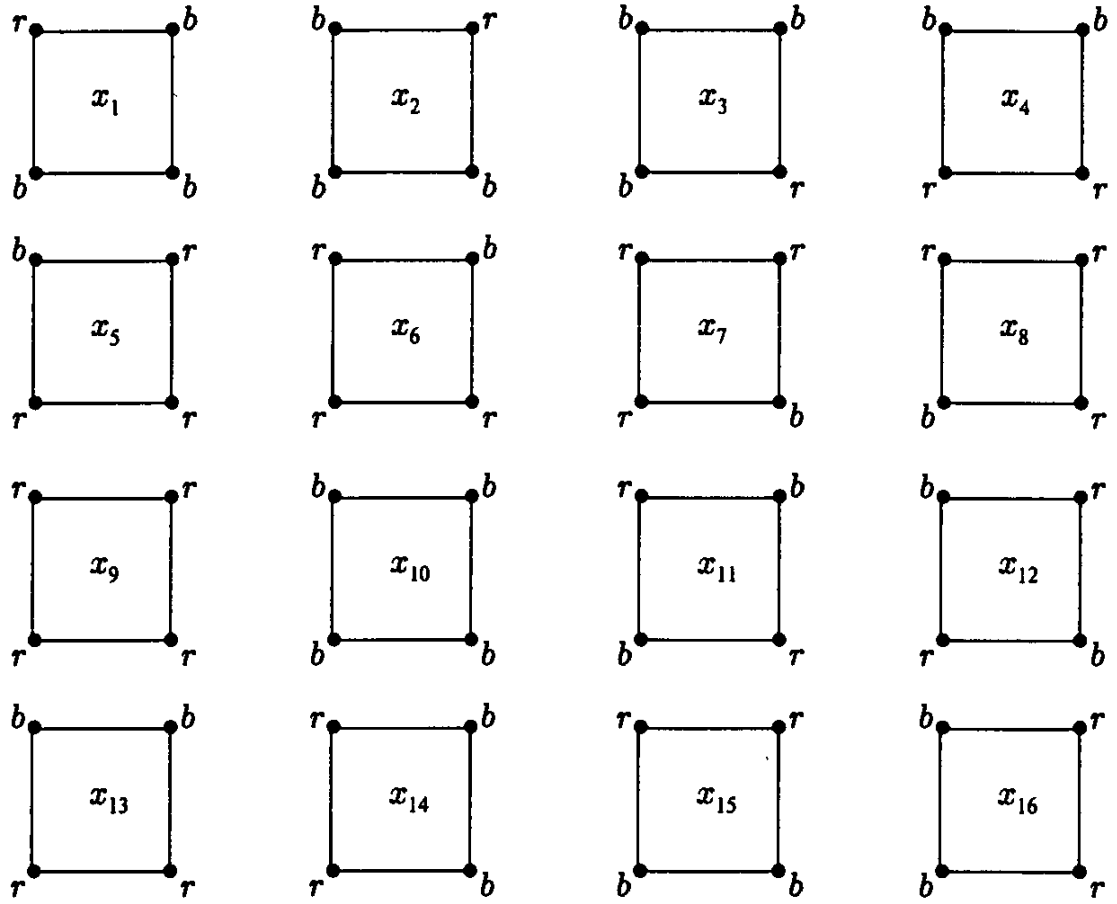


Рис. 1

Группа поворотов G имеет порядок 4 и является циклической. Она порождается поворотом на 90° по часовой стрелке, который обозначим как g. Тогда $(x_1)g = x_2$. Выпишем действие элементов группы на множестве 16 раскрасок:

$$e = g^0 = (x_1)(x_2)(x_3)(x_4)(x_5)(x_6)(x_7)(x_8)(x_9)(x_{10})(x_{11})(x_{12})(x_{13})(x_{14})(x_{15})(x_{16});$$

$$g^1 = (x_1, x_2, x_3, x_4)(x_5, x_6, x_7, x_8)(x_9)(x_{10})(x_{11}, x_{12})(x_{13}, x_{14}, x_{15}, x_{16});$$

$$g^2 = (x_1, x_3)(x_2, x_4)(x_5, x_7)(x_6, x_8)(x_9)(x_{10})(x_{11})(x_{12})(x_{13}, x_{15})(x_{14}, x_{16});$$

$$g^3 = (x_1, x_4, x_3, x_2)(x_5, x_8, x_7, x_6)(x_9)(x_{10})(x_{11}, x_{12})(x_{13}, x_{16}, x_{15}, x_{14}).$$

Подсчитав число стационарных точек и применив формулу Бернсайда для определения числа орбит, получим $(16+2+4+2)/4=6$. Эти 6 орбит, разумеется, снова будут циклами порождающего элемента g: $\{x_1,x_2,x_3,x_4\}$, $\{x_5,x_6,x_7,x_8\}$, $\{x_9\}$, $\{x_{10}\}$, $\{x_{11},x_{12}\}$, $\{x_{13},x_{14},x_{15},x_{16}\}$.

Может показаться, что непосредственное выписывание орбит позволяет найти их число быстрее, чем определение их числа с помощью подсчета числа стационарных точек и применения леммы Бернсайда. Однако лёгкость выписывания орбит в данном случае, с одной стороны, объясняется цикличностью группы G, а с другой стороны, это действие обозримо вследствие сравнительно небольшого числа раскрашенных квадратов, на которых действует группа G — всего 16. На множестве 729 раскрашенных кубов подобные действия, задаваемые вращениями, выписать в явном виде было бы значительно труднее. В то же время, нахождение числа стационарных раскрасок можно осуществить значительно проще, если рассмотреть действие группы G не на множестве 16 раскрашенных квадратов, а на множестве вершин квадрата, которых всего 4:

$$\begin{aligned} e=\ & g^0=(1)(2)(3)(4);\\ & g^1=(1,2,3,4);\\ & g^2=(1,3)(2,4);\\ & g^3=(1,4,3,2), \end{aligned}$$

и заметить, что некоторая раскраска является стационарной для данной подстановки в том и только в том случае, если все вершины каждого цикла подстановки выкрашены одним цветом. Таким образом, подстановка, состоящая из k циклов, имеет q^k стационарных раскрасок, где q — число цветов (в нашем случае $q=2$).

Сделанное замечание вместе с леммой Бернсайда позволяют найти число орбит как $(2^4+2^1+2^2+2^1)/4=6$. Это вычисление может быть представлено в изящной форме, если ввести *цикловой индекс* действия группы G на множестве вершин квадрата. Циклу длины p поставим в соответствие переменную c_p и каждую перестановку представим мономом, образованным произведением входящих в подстановку циклов: $e=g^0\to c_1^4$, $g^1\to c_4$, $g^2\to c_2^2$, $g^3\to c_4$. Просуммировав соответствующие подстановкам мономы и поделив полученный полином на порядок группы, получим *цикловой индекс* группы G от переменных c_1,c_2,c_4

$$Z_G(c_1,c_2,c_4)=\frac{1}{4}(c_1^4+c_2^2+2c_4).$$

Теперь для нахождений числа раскрасок достаточно подставить в цикловой индекс вместо каждой из переменных число используемых цветов $c_1 = c_2 = c_4 = 2$.

В общем случае цикловой индекс группы перестановок G степени n записывается как

$$Z_G(c_1, c_2, \ldots, c_n) = \frac{1}{|G|} \sum_{g \in G} c_1^{j_1} c_2^{j_2} \ldots c_n^{j_n}, \qquad (1)$$

где $(j_1, j_2, \ldots, j_n)$ — цикловой тип подстановки g.

Цикловой индекс позволяет не только получить полное число раскрасок, но также найти число раскрасок, использующих заданное число раз каждый из цветов. Для этого достаточно вместо каждой переменной c_p циклового индекса подставить выражение $(b^p + r^p)$ и привести полином к стандартному виду суммы произведений

$$\frac{1}{4}\left((b+r)^4 + (b^2+r^2)^2 + 2(b^4+r^4)\right) = b^4 + b^3 r + 2b^2 r^2 + br^3 + r^4.$$

Теперь коэффициент при мономе $b^i r^j$ показывает число раскрасок, использующих i раз синий и j раз красный цвет. Таким образом, имеется по одной одноцветной раскраске, по одной раскраске, использующей один раз один цвет и три раза другой, и две раскраски, использующие каждый из цветов дважды. Это легко проверить с помощью рис. 1.

Цикловой индекс был введён Дьердем Пойа (1887–1985) — известным венгерским математиком, с 1940 года проживавшим в США, соавтором ряда классических математических монографий, а также автором широко известных книг по методике преподавания математики. Ввиду основополагающей работы Пойа 1937 года (см. [10]), сильно продвинувшей теорию перечисления, методы перечисления, основанные на использовании циклового индекса, объединяют под общим названием теории Пойа. Обратимся, однако, к доказательству леммы Бернсайда.

Доказательство леммы Бернсайда. Пусть $\{x_1, x_2, \ldots, x_p\}$ — одна из орбит при действии группы G на множестве X. Пусть A_i — множество элементов $g \in G$ таких, что $(x_1)\, g = x_i$, $i = 1, 2, \ldots, p$. Покажем, что $A_i = G_{x_1} g$, где G_{x_1} — стабилизатор точки x_1, т. е. такая подгруппа группы G, элементы которой оставляют на месте точку x_1, а g — произвольный фиксированный элемент множества A_i. В самом деле, ясно что $G_{x_1} g \subseteq A_i$, так как $(x_1) G_{x_1} g = (x_1) g = x_i$. Докажем обратное включение. Пусть $g' \in A_i$, т. е. $(x_1) g' = x_i$. Тогда $g' g^{-1} \in G_{x_1}$, так как $(x_1) g' g^{-1} = x_1$. Поэтому $g' =$

$= g'\left(g^{-1}g\right) = \left(g'g^{-1}\right)g \in G_{x_1}g$. Этим доказывается, что $A_i \subseteq G_{x_1}g$, т. е. $A_i = G_{x_1}g$. Таким образом, $\{A_i,\ i = 1,2,\ldots,p\}$, — множество левых смежных классов по стабилизатору G_{x_1}. По теореме Лагранжа имеем

$$G = \bigcup_{i=1}^{p} A_i;$$

$$|G| = p\left|G_{x_1}\right|;$$

$$\left|G_{x_1}\right| = \frac{|G|}{p}.$$

Такую же мощность имеют и стабилизаторы остальных точек орбиты

$$\left|G_{x_i}\right| = \frac{|G|}{p},\ i = 1,2,\ldots,p.$$

Поэтому $\sum_{i=1}^{p}\left|G_{x_i}\right| = |G|$.

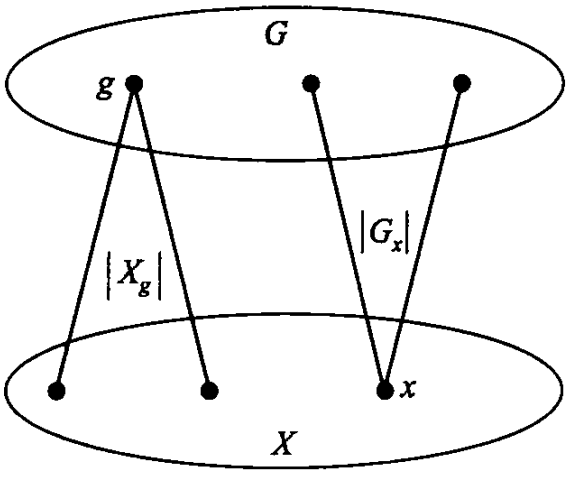


Рис. 2

Рассмотрим теперь двудольный граф (рис. 2), одну долю которого составляют элементы группы G, другую — точки множества X. Ребро $\{x, g\}$ проводится в том и только в том случае, если точка $x \in X$ является стационарной для элемента $g \in G$.

Подсчитаем число рёбер графа двумя способами: с верхней доли и с нижней доли. Подсчёт с верхней доли даёт $\sum_{g\in G}\left|X_g\right|$, а с нижней — $\sum_{x\in X}|G_x|$.

Если последнюю сумму разбить по орбитам, то вклад каждой орбиты по доказанному ранее будет равен $|G|$. Поэтому вся сумма будет равна $N|G|$, где N — число орбит. Имеем

$$N|G| = \sum_{g\in G}\left|X_g\right|.$$

Откуда

$$N = \frac{1}{|G|}\sum_{g\in G}\left|X_g\right|. \square$$

Вернёмся теперь к задаче о раскраске граней куба. С этой целью рассмотрим группу самосовмещающих вращений куба. Так как куб можно поставить на стол любой из 6 своих граней и для каждой такой постановки имеется ещё 4 различных самосовмещённых положения, то эта группа содержит $6 \cdot 4 = 24$ элемента, а именно:

1. Тождественное вращение.
2. 3 нетождественных вращения вокруг каждой из 3 осей четвёртого порядка, соединяющих середины противоположных граней, — итого 9 вращений.
3. 2 нетождественных вращения вокруг каждой из 4 осей третьего порядка, соединяющих противоположные вершины, — итого 8 вращений.
4. 1 нетождественное вращение вокруг каждой из 6 осей второго порядка, соединяющих середины противоположных ребер, — итого 6 вращений.

Выпишем цикловую структуру каждого из этих вращений как перестановки на множестве 6 граней.

1. Тождественное вращение оставляет каждую из 6 граней на месте, поэтому ей соответствует моном c_1^6 (шесть циклов длины 1).
2. Эти вращения по цикловой структуре разбиваются на 2 класса: вращения на 180° и вращения на ± 90°. Первые имеют цикловую структуру $c_1^2 c_2^2$ (два цикла длины 1 и два цикла длины 2), вторые — $c_1^2 c_4^1$ (два цикла длины 1 и один цикл длины 4).
3. Этим вращениям соответствуют мономы c_3^2 (два цикла длины 3).
4. Это вращение запишется как c_2^3 (три цикла длины 2).

Просуммировав мономы, выражающие циклические структуры всех перестановок, и поделив сумму на число элементов в группе, получим цикловой индекс группы

$$Z_G(c_1, c_2, c_3, c_4) = \frac{1}{24}\left(c_1^6 + 3c_1^2 c_2^2 + 6c_1^2 c_4^1 + 8c_3^2 + 6c_2^3\right).$$

Рецепт для нахождения числа раскрасок остаётся прежним. Достаточно в цикловой индекс вместо каждой из переменных подставить число используемых красок — в данном случае 3:

$$N = Z_G(3,3,...,3) = \frac{1}{24}\left(3^6 + 3 \cdot 3^2 \cdot 3^2 + 6 \cdot 3^2 \cdot 3 + 8 \cdot 3^2 + 6 \cdot 3^3\right) = 57.$$

В самом деле, как и прежде, для данной перестановки g граней куба раскраска будет стационарной в том и только в том случае, если грани, принадлежащие одному циклу, выкрашены в одинаковый цвет. Поэтому, если циклическая структура перестановки g есть $c_1^{i_1} c_2^{i_2} c_3^{i_3} c_4^{i_4}$, и используются

3 цвета, то она имеет $\left|X_g\right| = 3^{i_1} \cdot 3^{i_2} \cdot 3^{i_3} \cdot 3^{i_4}$ стационарных раскрасок, и рецепт является прямым следствием леммы Бернсайда.

Для того чтобы найти число раскрасок, использующих i раз белый, j раз синий и k раз красный цвет, нужно в цикловой индекс вместо каждой переменной c_l, $l = 1, 2, ..., 6$, подставить выражение $\left(w^l + b^l + r^l\right)$ и, раскрыв скобки в выражении

$$Z_G\left((w+b+r),\left(w^2+b^2+r^2\right),\left(w^3+b^3+r^3\right),\left(w^4+b^4+r^4\right)\right),$$

найти коэффициент при $w^i b^j r^k$. Этот рецепт и составляет основное содержание теории Пойа. Обоснуем его.

Будем рассматривать действие группы вращений лишь на раскрасках типа $w^i b^j r^k$ и подсчитаем число орбит $N_{w^i b^j r^k}$ на этом множестве. Лемма Бернсайда даёт

$$N_{w^i b^j r^k} = \frac{1}{|G|} \sum_{g \in G} \left| X_g^{w^i b^j r^k} \right|,$$

где $X_g^{w^i b^j r^k}$ – множество раскрасок типа $w^i b^j r^k$, не изменяющихся при вращении g. Теперь для обоснования рецепта Пойа остаётся лишь заметить, что для подсчёта числа стационарных раскрасок у вращения с циклической структурой $c_1^{i_1} c_2^{i_2} c_3^{i_3} c_4^{i_4}$ нужно вместо каждой переменной c_p подставить выражение $(w^p + b^p + r^p)$ и, раскрыв скобки, найти коэффициент при мономе $w^i b^j r^k$. Это является прямым следствием того, что для стационарности раскраски необходимо и достаточно, чтобы все грани цикла были выкрашены одним цветом. Каждый цикл длины p в стационарной раскраске даёт p одноцветных граней, и выбор переменной при раскрытии скобки $(w^p + b^p + r^p)$ в процессе перемножения скобок соответствует выбору цвета при раскраске этих граней.

В качестве примера найдём число раскрасок, использующих каждый цвет дважды. Коэффициент, получающийся при $w^2 b^2 r^2$ после раскрытия скобок в выражении

$$\frac{1}{24}\Big((w+b+r)^6 + 3(w+b+r)^2 (w^2+b^2+r^2)^2 +$$

$$+ 6(w+b+r)^2 (w^4+b^4+r^4) + 8(w^3+b^3+r^3)^2 + 6(w^2+b^2+r^2)^3\Big),$$

равен $\frac{1}{24}\left(\frac{6!}{2!2!2!}+3(1\cdot2+1\cdot2+1\cdot2)+6\frac{3!}{1!1!1!}\right)=6$, что и даёт число искомых раскрасок. (Читателю предлагается самостоятельно найти эти раскраски)

В качестве еще одного примера на применение теоремы Пойа рассмотрим следующую задачу.

> Сколько различных ожерелий из 7 бусинок можно составить, используя 3 красных и 4 синих бусинки?

Ожерелья, получающиеся передвижением бусинок по нитке ожерелья, а также переворотом ожерелья, считаются одинаковыми. Поэтому с математической точки зрения задача эквивалентна числу раскрасок вершин правильного 7-угольника, использующих 3 раза красный цвет и 4 раза синий.

Группа самосовмещений 7-угольника состоит из 7 поворотов на углы 0°, 360°/7, 2·360°/7, 3·360°/7, 4·360°/7, 5·360°/7, 6·360°/7 и 7 переворотов относительно прямых, соединяющих вершины с серединами противоположных сторон (рис. 3).

Поворот на нулевой угол как тождественное преобразование имеет цикловую структуру c_1^7, а действия остальных шести поворотов на множестве вершин семиугольника вследствие простоты числа 7 являются циклическими, т. е. имеют цикловую структуру c_7. Каждый переворот оставляет одну из вершин на месте, а остальные 6 разбивает на циклы длины 2, т. е. имеет цикловую структуру $c_1^1c_2^3$. Поэтому цикловой индекс группы самосовмещений 7-угольника есть

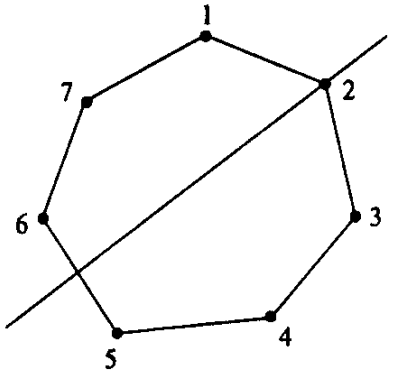


Рис. 3

$$Z_G(c_1,c_2,c_7)=\frac{1}{14}(c_1^7+6c_7^1+7c_1c_2^3).$$

Теперь для нахождения числа раскрасок, использующих 3 раза красный цвет и 4 раза синий, нужно, согласно рецепту Пойа, раскрыв выражение

$$\frac{1}{14}\left((r+b)^7+6(r^7+b^7)+7(r+b)(r^2+b^2)^3\right),$$

найти коэффициент при r^3b^4. Этот коэффициент, как нетрудно заметить, равен

$$\frac{1}{14}\left(C_7^3+7C_3^1\right)=4.$$

Таким образом, существует 4 ожерелья заданного типа. Вот они (рис. 4):

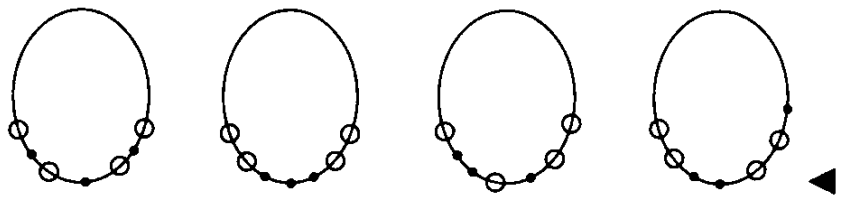


Рис. 4

Если бы число бусинок не было простым числом, то не все повороты были бы циклическими перестановками, некоторые из них распадались бы на несколько циклов одинаковой длины. Дадим общую формулу для циклового индекса группы симметрий правильного n-угольника (группы диэдра D_n). Эта группа порядка $2n$ состоит из n вращений на углы $2\pi k/n$, $k = 0,1, \ldots, n-1$, и n отражений. Рассмотрим сначала вращения.

Подгруппа вращений G является циклической группой, порождаемой вращением на угол $2\pi/n$. Обозначив это вращение буквой g, имеем $G = \{e = g^0, g^1, g^2, \ldots, g^{n-1}\}$. Если порождаемая вращением перестановка распадается на n/d циклов длины d, то она имеет цикловую структуру $c_d^{n/d}$. Для любого d, делящего n, такой цикловой структурой обладает элемент $g^{n/d}$, а также все элементы вида $(g^{n/d})^k$, где $(k, d) = 1$, и только они. Поэтому для каждого d, делящего n, имеется $\varphi(d)$ перестановок с цикловой структурой $c_d^{n/d}$, где $\varphi(d)$ — функция Эйлера.

Обратившись к отражениям, отдельно рассмотрим случай чётного и нечётного n. В случае нечётного n все n отражений имеют цикловую структуру $c_1 c_2^{(n-1)/2}$, так как каждая ось симметрии проходит через одну из вершин правильного n-угольника. В случае же чётного n половина отражений имеет цикловую структуру $c_1^2 c_2^{(n-2)/2}$, когда ось проходит через две противоположные вершины, и половина — $c_2^{n/2}$, когда ось не пересекает вершин.

Окончательно получаем, что цикловой индекс диэдральной группы D_n равен

$$Z_{D_n}(c_1, \ldots, c_n) = \frac{1}{2n}\left(\sum_{d|n} \varphi(d) c_d^{n/d} + n c_1 c_2^{(n-1)/2}\right)$$

в случае нечётного n и равен

$$Z_{D_n}(c_1, \ldots, c_n) = \frac{1}{2n}\left(\sum_{d|n} \varphi(d) c_d^{n/d} + \frac{n}{2} c_1^2 c_2^{(n-2)/2} + \frac{n}{2} c_2^{n/2}\right)$$

в случае чётного n.

В качестве примера выпишем цикловой индекс группы D_6. Значения функции Эйлера для четырёх делителей числа 6 равны $\varphi(1) = 1$, $\varphi(2) = 1$, $\varphi(3) = 2$, $\varphi(6) = 2$, и цикловой индекс принимает вид

$$Z_{D_6}(c_1, \ldots, c_6) = \frac{1}{12}\left(c_1^6 + c_2^3 + 2c_3^2 + 2c_6 + 3c_1^2 c_2^2 + 3c_2^3\right).$$

Пойа развивал свою теорию, имея в виду её конкретное применение к перечислению химических изомеров органических молекул заданной структуры. В качестве простейшего примера рассмотрим молекулы вида, показанного на рис. 5.

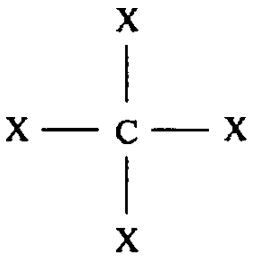
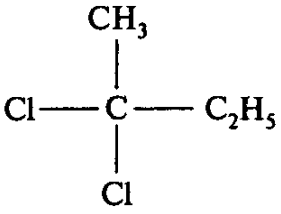
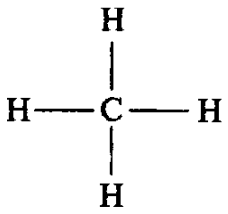


Рис. 5



дихлорбутан



метан

Рис. 6

Здесь C — атом углерода, чёрточками обозначены валентные связи, а на месте буквы «x» находится одно из следующих соединений: CH_3 (метил), C_2H_5 (этил), H (водород), Cl (хлор). Например, как на рис. 6.

Математической моделью подобной молекулы будем считать тетраэдр, в центре которого находится атом углерода, а в каждой вершине — одно из перечисленных соединений.

Группа самосовмещений тетраэдра состоит из 12 вращений, цикловая структура действия которых на множестве вершин имеет вид:

1) тождественное вращение — c_1^4;
2) восемь нетождественных вращений вокруг осей, проходящих через вершины и центры противолежащей грани, — c_1c_3;
3) три вращения вокруг осей, проходящих через середины противолежащих рёбер, — c_2^2.

Цикловой индекс группы

$$Z_G = \frac{1}{12}(c_1^4 + 8c_1c_3 + 3c_2^2).$$

Подставив в цикловой вместо каждой переменной значение 4, получим, что полное число различных молекул равно 36. Заметим, что это число на единицу превышает $\overline{C}_4^4 = 35$, т. е. имеются две молекулы, одинаковые по составу, но различающиеся между собой. Это молекулы, содержащие все 4 используемых соединения, но имеющие различную ориентацию. Являясь зеркально симметричными, они не могут быть совмещены с помощью вращения.

Обобщим теперь рассмотренные выше задачи о числе раскрасок куба, числе ожерелий и числе изомеров, вписав их в единую математическую модель. Пусть функция $f : X \to Y$ отображает конечное множество X (область определения) в конечное множество Y (область значений). Если $|X| = n$, $|Y| = m$, то существует m^n таких функций. Пусть, далее, на мно-

жестве X действует некоторая группа подстановок G. Назовём две функции f_1 и f_2 эквивалентными, если существует такая подстановка $g \in G$, что $f_1(x) = f_2((x)g)$ для всех $x \in X$ (напомним, что в соответствии с принятым соглашением переставляемый элемент пишется в скобках слева от подстановки). Задача состоит в подсчёте числа классов эквивалентности.

Если группа G состоит лишь из тождественной подстановки, то число классов эквивалентности совпадает с числом функций и равно m^n. В другом крайнем случае, когда группа G является симметрической группой S_n, число классов эквивалентности становится равным $\bar{C}_m^n$ (см. табл. 1.1).

Теория Пойа позволяет с помощью циклового индекса решать подобного рода задачи для произвольной группы G, находя как общее число классов эквивалентности, так и число классов с заданным числом принимаемых функцией f различных значений из множества Y. В случае раскрасок куба в качестве множества X выступало множество граней куба, в качестве множества Y — множество используемых цветов, а группой G была группа самосовмещающих вращений куба. В задаче с ожерельем множеством X являлось множество вершин правильного n-угольника, множеством Y — множество типов бусинок, а в качестве группы G выступала диэдральная группа D_n.

В заключении выпишем в явном виде цикловой индекс симметрической группы S_n. Используя (1.8), получаем

$$Z_{S_n}(c_1, c_2, \ldots, c_n) = \frac{1}{n!} \sum_{\substack{j_1, j_2, \ldots, j_n \geq 0 \\ 1j_1+2j_2+\ldots+nj_n=n}} \frac{n!}{\prod_{i=1}^{n} i^{j_i} j_i!} c_1^{j_1} c_2^{j_2} \ldots c_n^{j_n}. \qquad (2)$$

В качестве примера приведём цикловой индекс для S_4

$$Z_{S_4}(c_1, c_2, c_3, c_4) = \frac{1}{24}(c_1^4 + 6c_1^2 c_2 + 8c_1 c_3 + 3c_2^2 + 6c_4). \qquad (3)$$

Вопросы для самопроверки

1. Цикловой индекс группы самосовмещающих вращений куба, действующей на множестве его вершин, равен

   а) $\frac{1}{24}\left(c_1^8 + 6c_4^2 + 9c_2^4 + 8c_1^2 c_3^2\right)$; б) $\frac{1}{24}\left(c_1^8 + 8c_1^2 c_3^2 + 15c_2^4\right)$;

   в) $\frac{1}{24}\left(c_1^8 + 16c_1^2 c_3^2 + 7c_4^2\right)$.

2. Число различных раскрасок вершин куба в 2 цвета с учетом самосовмещений равно

   а) 21; б) 22; в) 23.

3. Число различных раскрасок вершин куба в 2 цвета, использующих 3 раза красный и 5 раз синий цвет, с учетом самосовмещений равно
а) 2; б) 3; в) 4.

Ответы: 1 — а, 2 — в, 3 — б.

1.7. Асимптотические оценки. Формула Стирлинга

Асимптотические обозначения. При вычислении некоторого комбинаторного числа $f(n)$, зависящего от натурального параметра $n = 1, 2, \ldots$, часто оказывается, что с ростом n стремительно нарастает как $f(n)$, так и сложность вычислений этой функции, а формула, выражающая $f(n)$ как функцию от n в общем случае является чрезвычайно сложной и труднообозримой. В этом случае имеет смысл оценить $f(n)$ при больших значениях n с помощью другой, более простой функции $g(n)$, близкой в некотором смысле к $f(n)$. Подобные оценки называют *асимптотическими* и часто используют в дискретной математики. Наиболее часто в качестве подобной оценки выступает так называемое *асимптотическое равенство*, записываемое как

$$f(n) \sim g(n), \qquad n \to \infty,$$

которое означает, что $\lim\limits_{n\to\infty} \dfrac{f(n)}{g(n)} = 1$.

Примерами асимптотических равенств являются:

$$e^n + n \sim e^n, \quad n \to \infty; \quad n + \ln n \sim n, \quad n \to \infty \text{ и т. д.}$$

Асимптотическое равенство является отношением эквивалентности, разбивающим функции на классы эквивалентности. В частности, в один класс попадают функции

$$n^2 \sim n^2 + 2 \sim n^2 + 5n\sqrt{n} + 1 \sim n^2 + \ln^3 n, \quad n \to \infty.$$

Часто оказывается нужным оценить скорость роста величины $f(n)$ сверху. В этом случае оказывается полезным обозначение

$$f(n) = O\big(g(n)\big),$$

которое читается: «$f(n)$ есть *„о"-большое от* $g(n)$», и означает, что существует некоторая константа C такая, что $|f(n)| \le C|g(n)|$ для всех $n = 1, 2, \ldots$. Например, $5n^2 + 2n + 10 = O(n^2)$. Запись $f(n) = O(1)$ обозначает ограниченность $|f(n)|$. Если $f(n) = O\big(g(n)\big)$, то говорят также, что $f(n)$ *по порядку* не превосходит $g(n)$. А если одновременно имеет место

$f(n)=O(g(n))$ и $g(n)=O(f(n))$, то говорят, что $f(n)$ и $g(n)$ являются величинами *одного порядка*, и пишут $f(n)=\Theta(g(n))$.

Напомним здесь также другую, широко используемую в математическом анализе запись

$$f(x)=o(g(x)),\quad x\to\infty,$$

которая читается: «$f(x)$ есть *„o"-малое от* $g(x)$» и означает, что $\lim\limits_{x\to\infty}\dfrac{f(x)}{g(x)}=0$. Например, $x=o(e^x)$, $x\to\infty$. Запись $f(x)=o(1)$ означает, что $\lim\limits_{x\to\infty}f(x)=0$, т. е. $f(x)$ является бесконечно малой величиной. Заметим, что асимптотическое равенство $f(n)\sim g(n)$, $n\to\infty$, может быть записано также как $f(n)=g(n)(1+o(1))$, $n\to\infty$.

Асимптотическое неравенство

$$f(n)\lesssim g(n)$$

означает, что для некоторого $o(1)$ имеет место неравенство

$$f(n)\le g(n)(1+o(1)).$$

Заметим, что выполнение обоих соотношений $f(n)\lesssim g(n)$ и $g(n)\lesssim f(n)$ означает, что $f(n)\sim g(n)$, $n\to\infty$.

Формула Стирлинга. Функция $n!$, несомненно, занимает центральное место в комбинаторике. Однако, по своему определению, для больших n она является достаточно сложной, неудобной для оценок конструкцией, не только точное значение, но и порядок величины которой непосредственно не виден. Поэтому при больших n как в теоретических исследованиях, так и при практических вычислениях для $n!$ используются оценки, имеющие более простой и удобный вид. Нередко используется следующая несложная универсальная нижняя оценка

$$n!>\left(\frac{n}{e}\right)^n. \tag{1}$$

Её легко пояснить тем, кто знаком с рядами Тейлора—Маклорена. В самом деле,

$$e^x=1+\frac{x}{1!}+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\frac{x^5}{5!}+\dots.$$

Поэтому

$$e^n=1+\frac{n}{1!}+\frac{n^2}{2!}+\frac{n^3}{3!}+\frac{n^4}{4!}+\frac{n^5}{5!}+\dots.$$

Оставляя в разложении лишь один n-й член, получаем

$$e^n > \frac{n^n}{n!},$$

откуда и следует (1).

Более точной, однако, является часто используемая при больших n для оценки факториалов *асимптотическая формула Стирлинга*

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n, \quad n \to \infty. \tag{2}$$

Эта знаменитая асимптотическая формула была в 1730 году получена шотландцем Джеймсом Стирлингом (1692–1770) и Абрахамом де Муавром, использовавшим её для доказательства предельной теоремы теории вероятностей, носящей имя Муавра—Лапласа.

Хотя разность между левой и правой частями формулы (2) неограниченно возрастает, относительная погрешность быстро стремится к нулю. Так, для $5! = 120$ правая часть (2) даёт 118,02, а для $10! = 3\,628\,800$ – 3 598 600. В первом случае относительная погрешность составляет около 2 %, а во втором — уже менее 1 %.

Чтобы продемонстрировать возможность применения формулы Стирлинга, рассмотрим следующую известную задачу. Пусть имеется некоторое число k людей, скажем, $20 \le k \le 50$. Какова вероятность, что ни у кого из них не совпадают дни рождения? На первый взгляд может показаться, что даже при $k = 50$ вероятность различия всех дней рождения будет близка к единице, так количество людей составляет лишь незначительную часть от числа дней в году $n = 365$. Произведём, однако, необходимые вычисления. Искомая вероятность есть

$$\frac{A_n^k}{n^n} = \frac{n!}{(n-k)!n^n}.$$

Считая, что n и $n-k$ достаточно велики, а k не слишком велико, так что $\frac{k}{n-k} << 1$, с помощью формулы Стирлинга получаем

$$\frac{n!}{(n-k)!n^n} \approx \frac{\sqrt{2\pi n}\, n^n e^{-n}}{\sqrt{2\pi(n-k)}\,(n-k)^{n-k} e^{-(n-k)} n^k} = \left(\frac{n}{n-k}\right)^{n-k+1/2} e^{-k} =$$

$$= e^{(n-k+\frac{1}{2})\ln(1+\frac{k}{n-k})-k}.$$

Ограничиваясь в разложении логарифма двумя первыми членами $\ln(1+x) \approx x - x^2/2$, далее имеем

$$e^{(n-k+\frac{1}{2})\ln(1+\frac{k}{n-k})-k} \approx e^{(n-k+\frac{1}{2})(\frac{k}{n-k}-\frac{1}{2}(\frac{k}{n-k})^2)-k} \approx e^{\frac{1}{2}\frac{k}{n-k}-\frac{1}{2}\frac{k^2}{n-k}} = e^{\frac{k(1-k)}{2(n-k)}}.$$

Это даёт несколько неожиданный результат. При $k = 20$ искомая вероятность близка к $0{,}6$, а при $k = 50$ — к $0{,}02$. Событие, имеющее вероятность $0{,}02$, часто считают практически невозможным!

Доказательство формулы Стирлинга проведём, следуя [20]. Имеем

$$\ln n! = \ln 1 + \ln 2 + \ldots + \ln n .$$

Так как $\ln x$ — монотонная функция, справедливо двойное неравенство

$$\int_{k-1}^{k} \ln x\,dx < \ln k < \int_{k}^{k+1} \ln x\,dx .$$

Суммируя по $k = 1, 2, \ldots, n$, получаем

$$\int_{0}^{n} \ln x dx < \ln n! < \int_{1}^{n+1} \ln x dx ,$$

$$n\ln n - n < \ln n! < (n+1)\ln(n+1) - n .$$

Это двойное неравенство подсказывает приближать $\ln n!$ величиной, близкой к среднему арифметическому крайних членов неравенства. Простейшей такой величиной является $(n+1/2)\ln n - n$. Взяв эту величину за основу, рассмотрим разность

$$d_n = \ln n! - \left(n + \frac{1}{2}\right)\ln n + n . \tag{3}$$

Заметим, что

$$d_n - d_{n+1} = \left(n + \frac{1}{2}\right)\ln\frac{n+1}{n} - 1 = \frac{1}{2}(2n+1)\ln\frac{1+1/(2n+1)}{1-1/(2n+1)} - 1 . \tag{4}$$

Для дальнейшего анализа последовательности $\{d_n\}$ нам потребуется разложение в ряд функции $\ln\dfrac{1+x}{1-x}$. С этой целью воспользуемся формулой для суммы бесконечно убывающей геометрической прогрессии

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots .$$

Проинтегрировав её почленно, получим

$$-\ln(1-x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \ldots . \tag{5}$$

Подставляя в (5) $(-x)$ вместо x, имеем

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots . \tag{6}$$

Складывая (5) и (6), получаем

$$\ln\frac{1+x}{1-x}=2\left(x+\frac{x^3}{3}+\frac{x^5}{5}+\ldots\right). \tag{7}$$

Теперь, возвращаясь к (4) и используя (7), имеем

$$d_n-d_{n+1}=\frac{1}{3(2n+1)^2}+\frac{1}{5(2n+1)^4}+\frac{1}{7(2n+1)^6}+\ldots<$$

$$<\frac{1}{3(2n+1)^2}\left(1+\frac{1}{(2n+1)^2}+\frac{1}{(2n+1)^4}+\ldots\right)=$$

$$=\frac{1}{3(2n+1)^2}\,\frac{1}{1-(2n+1)^{-2}}=\frac{1}{12n(n+1)}=\frac{1}{12n}-\frac{1}{12(n+1)}. \tag{8}$$

Таким образом, получаем двойное неравенство

$$0<d_n-d_{n+1}<\frac{1}{12n}-\frac{1}{12(n+1)}\,. \tag{9}$$

Из (9) следует, что $\{d_n\}$ является монотонно убывающей последовательностью. Кроме того, из (9) вытекает, что

$$d_n-\frac{1}{12n}<d_{n+1}-\frac{1}{12(n+1)},$$

т. е. последовательность $\left\{d_n-\frac{1}{12n}\right\}$ монотонно возрастает, откуда следует, что $\left\{d_n-\frac{1}{12n}\right\}$, а следовательно, и $\{d_n\}$ ограничены снизу. Поэтому последовательность $\{d_n\}$ сходится. Пусть

$$\lim_{n\to\infty}d_n=C\,.$$

Тогда, ввиду (3), имеем

$$\lim_{n\to\infty}\left(\ln n!-\left(n+\frac{1}{2}\right)\ln n+n\right)=C,$$

что при потенцировании даёт

$$\lim_{n\to\infty}\frac{n!}{\sqrt{n}\left(\frac{n}{e}\right)^n}=e^C,$$

т. е.

$$n!\sim e^C\sqrt{n}\left(\frac{n}{e}\right)^n,\qquad n\to\infty. \tag{10}$$

Покажем теперь, что $e^C = \sqrt{2\pi}$. Пусть $e^C = \kappa$. Рассмотрим сумму

$$\sum_{i=0}^{2n} C_{2n}^i = 2^{2n}. \tag{11}$$

Максимальный член в этой сумме есть C_{2n}^n. Используя (10), имеем

$$C_{2n}^n \sim \frac{(2n)!}{n!n!} \sim \frac{\kappa\sqrt{2n}\,(2n)^{2n}\,e^{-2n}}{\kappa\sqrt{n}\,n^n e^{-n}\,\kappa\sqrt{n}\,n^n e^{-n}} = \frac{1}{\kappa}\sqrt{\frac{2}{n}}\,2^{2n}, \quad n\to\infty. \tag{12}$$

Сравним теперь с C_{2n}^n остальные члены суммы (11), которые будет удобно записывать, как $\{C_{2n}^{n+k}\}$, где k изменяется от $-n$ до n.

$$C_{2n}^{n+k} = C_{2n}^n \frac{(n-k+1)\cdot\ldots\cdot(n-1)\cdot n}{(n+1)\cdot(n+2)\cdot\ldots\cdot(n+k)} = C_{2n}^n \frac{(1-\frac{k-1}{n})\cdot\ldots\cdot(1-\frac{1}{n})\cdot 1}{(1+\frac{1}{n})\cdot(1+\frac{1}{n})\cdot\ldots\cdot(n+\frac{k}{n})}.$$

Устремляя $n\to\infty$ и используя разложение $e^{i/n} = 1+\frac{i}{n}+O(\frac{i}{n})^2$, имеем

$$C_{2n}^{n+k} = C_{2n}^n \exp\left\{-\frac{1}{n}\left(2[1+2+3+\ldots+(k-1)]+k\right)+O\left(k^3/n^2\right)\right\} =$$

$$= C_{2n}^n \exp\left\{-\frac{k^2}{n}+O\left(\frac{k^3}{n^2}\right)\right\}. \tag{13}$$

Пусть теперь k возрастает с ростом n так, что $\frac{k^2}{n}\to\infty$, а $\frac{k^3}{n^2}\to 0$, например, $k = [n^{3/5}]$. Тогда $C_{2n}^k \sim C_{2n}^n e^{-k^2/n}$ и $\sum_{i=-k}^{k} C_{2n}^{n+i} \sim \sum_{i=-n}^{n} C_{2n}^{n+i} = 2^{2n}$. Это позволяет записать следующую цепочку асимптотических равенств

$$2^{2n} = \sum_{i=-n}^{n} C_{2n}^{n+i} \sim \sum_{i=-k}^{k} C_{2n}^{n+i} \sim C_{2n}^n \sum_{i=-k}^{k} e^{-i^2/n} \sim C_{2n}^n \int_{-k}^{k} e^{-\xi^2/n}d\xi \sim$$

$$\sim C_{2n}^n \int_{-\infty}^{\infty} e^{-\xi^2/n}d\xi \sim C_{2n}^n\sqrt{n}\int_{-\infty}^{\infty} e^{-x^2}dx \sim \frac{1}{\kappa}\sqrt{\frac{2}{n}}\,2^{2n}\sqrt{n}\int_{-\infty}^{\infty} e^{-x^2}dx =$$

$$= \frac{\sqrt{2}}{\kappa}2^{2n}\int_{-\infty}^{\infty} e^{-x^2}dx = \frac{\sqrt{2}}{\kappa}2^{2n}\sqrt{\pi}.$$

(Напомним, что $\int_{-\infty}^{\infty} e^{-x^2}dx = \sqrt{\pi}$ есть хорошо известный в математическом анализе интеграл, часто называемый интегралом Пуассона)

Отсюда получаем, что $\kappa = \sqrt{2\pi}$.

Формула Стирлинга доказана! □

Заметим, что, продолжая исследование в духе уже проведённых рассуждений (см. [20]), можно получить и более точные оценки для факториала

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n}}.$$

Обобщение факториала. Как для теоретических целей, так и для практики весьма желательно было бы найти аналитическую функцию, определённую на всей положительной полуоси и совпадающую в её целых точках с $n!$. Такая функция, интерполирующая факториал для дробных значений аргумента, была в 1729 году предложена Эйлером и вскоре стала одной из самых известных специальных функций.

Для натурального n интегрирование выражения $\int_0^\infty x^n e^{-x}dx$ по частям даёт

$$\int_0^\infty x^n e^{-x}dx = -\int_0^\infty x^n de^{-x} = -x^n e^{-x}\Big|_0^\infty + n\int_0^\infty x^{n-1}e^{-x}dx = n\int_0^\infty x^{n-1}e^{-x}dx.$$

Повторяя этот процесс n раз, получаем

$$\int_0^\infty x^n e^{-x}dx = n(n-1)(n-2)\ldots 2\int_0^\infty e^{-x}dx = n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot 2\cdot 1 = n!$$

Тем самым для $n!$ получено простое аналитическое выражение в виде интеграла, зависящего от n как от параметра:

$$n! = \int_0^\infty x^n e^{-x}dx \quad (n = 0,1,2,\ldots). \tag{14}$$

Но интеграл в правой части равенства (14) определён не только для целых положительных n, но для всех действительных $n > -1$. Соотношение (14) позволяет, таким образом, расширить область определения факториала. График определённой для $z \geq -1$ функции $z! = \int_0^\infty x^z e^{-x}dx$ представлен на рис. 1.

Для определённой таким образом для $z > -1$ функции $z!$ сохраняются основные свойства факториала:

$$z! = z(z-1)!;$$

$$z! \sim \sqrt{2\pi z}\, z^z e^{-z}, \quad z \to \infty;$$

$$z! = \sqrt{2\pi z}\, z^z e^{-z+\frac{\theta}{12z}}, \quad 0 < \theta < 1.$$

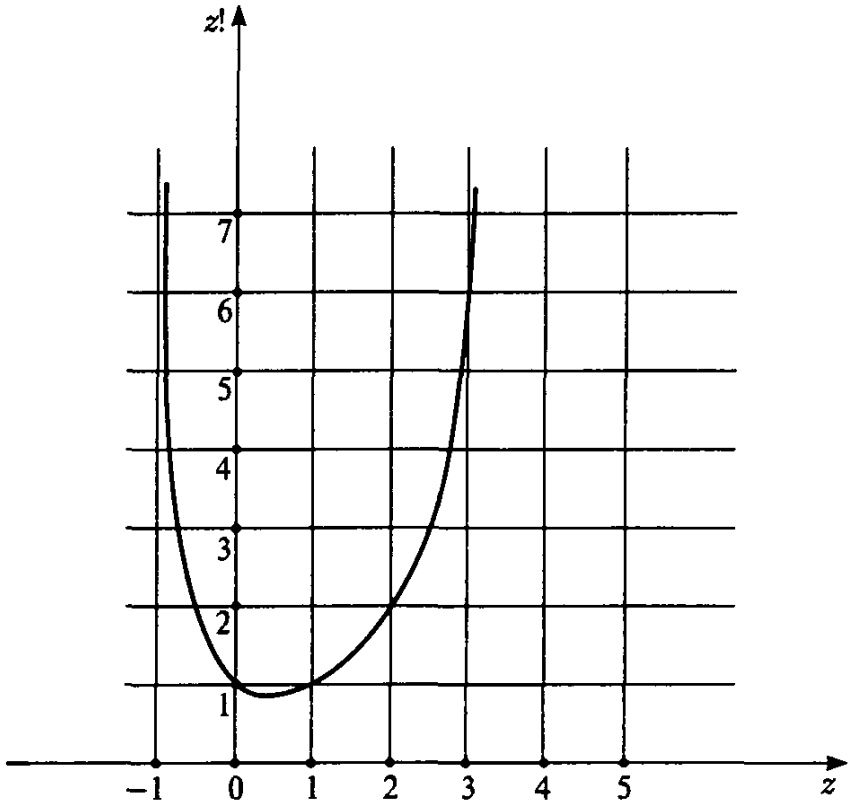


Рис. 1

Можно показать, что $(1/2)! = \sqrt{\pi}/2$, а минимальное значение функции $z!$ достигается при $z \approx 0{,}46163$ и равно $\approx 0{,}88560$.

Хотя запись $z!$ для действительного $z > -1$ и является вполне корректной, в случае дробного z в обозначениях обычно используют *гамма-функцию*

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx,$$

называемую также эйлеровым интегралом второго рода. Связь между факториалом и гамма-функцией выражается равенством

$$z! = \Gamma(z+1) = z\,\Gamma(z).$$

Оценки для биномиальных коэффициентов и их сумм. В задачах комбинаторного анализа биномиальным коэффициентам и связанным с ними оценкам принадлежит исключительно важная роль. Эти оценки часто носят асимптотический характер и опираются на формулу Стирлинга. Они входят в обязательный технический арсенал занимающихся комбинаторный анализом.

Во многих случаях оказываются полезными следующие простые универсальные оценки для величины биномиальных коэффициентов:

$$\left(\frac{n}{k}\right)^k < C_n^k < \left(\frac{en}{k}\right)^k. \qquad (15)$$

Имеем

$$C_n^k = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!} < \frac{n^k}{k!} < \frac{n^k}{(k/e)^k} = \left(\frac{en}{k}\right)^k,$$

где для оценки $k!$ было использовано неравенство (1).

Нижнюю же оценку в (15) даёт следующая выкладка

$$C_n^k = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k(k-1)(k-2)\quad\ldots\quad 1} = \frac{n}{k}\frac{n-1}{k-1}\frac{n-2}{k-2}\cdots\frac{n-k+1}{1} > \left(\frac{n}{k}\right)^k,$$

так как $\frac{n-i}{k-i} > \frac{n}{k}$ для $i = 1, 2, \ldots, k-1$ при $k < n$.

Более точные асимптотические оценки можно получить с помощью формулы Стирлинга. Биномиальный коэффициент C_n^k при фиксированном n достигает максимального значения при $k = [n/2]$. Согласно (12)

$$C_n^{[n/2]} \sim \sqrt{\frac{2}{\pi}}\,\frac{2^n}{\sqrt{n}}, \quad n \to \infty. \tag{16}$$

Часто весьма важным оказывается умение оценивать биномиальные коэффициенты C_n^k и их суммы в средней части ряда, когда $k \sim n/2$. Как следует из (13)

$$C_n^{[n/2]+k} \sim C_n^{[n/2]} e^{-2k^2/n}, \quad n \to \infty, \ k^3/n^2 \to 0, \tag{17}$$

Найдём теперь асимптотику для сумм биномиальных коэффициентов в средней части ряда, а именно сумм вида $\sum_{-t\sqrt{n}}^{t\sqrt{n}} C_n^{[n/2]+i}$, где t — константа. Имеем

$$\sum_{i=-t\sqrt{n}}^{t\sqrt{n}} C_n^{[n/2]+i} \sim C_n^{[n/2]} \sum_{i=-t\sqrt{n}}^{t\sqrt{n}} e^{-2i^2/n} \sim \sqrt{\frac{2}{\pi}}\,\frac{2^n}{\sqrt{n}} \int_{-t\sqrt{n}}^{t\sqrt{n}} e^{-2\xi^2/n} d\xi =$$

$$= \frac{2^n}{\sqrt{2\pi}} \int_{-2t}^{2t} e^{-\frac{x^2}{2}} dx = 2^n \cdot 2\Phi(2t), \tag{18}$$

где $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{x^2}{2}} dx$ — хорошо известная, табулированная функция, которую часто называют функцией Лапласа или интегралом вероятностей, $\lim_{z\to\infty} \Phi(z) = 1/2$.

Асимптотическое равенство (18) эквивалентно знаменитой предельной теореме Муавра—Лапласа в теории вероятностей о числе успехов в серии из n испытаний Бернулли в симметричном случае, когда $p = q = 1/2$

и все 2^n последовательностей из успехов и неудач длины n имеют одну и ту же вероятность $1/2^n$. Согласно (4.3) и (18)

$$\mathrm{P}_n\left\{[n/2]-t\sqrt{n}\le k\le[n/2]+t\sqrt{n}\right\}=\frac{1}{2^n}\sum_{k=[n/2]-t\sqrt{n}}^{[n/2]+t\sqrt{n}}C_n^k\sim 2\Phi(2t),\ n\to\infty. \quad (19)$$

Если в биномиальном коэффициенте C_n^k параметр k растёт существенно медленнее, чем n, то справедливо следующее асимптотическое равенство

$$C_n^k\sim\frac{n^k}{k!}e^{\frac{k^2}{2n}},\quad n\to\infty,\quad k=o(n^{2/3}). \quad (20)$$

Имеем

$$C_n^k=\frac{n(n-1)(n-2)\ldots(n-k+1)}{k!}=\frac{n^k}{k!}\cdot 1\cdot\left(1-\frac{1}{n}\right)\cdot\left(1-\frac{2}{n}\right)\cdot\ldots\cdot\left(1-\frac{k-1}{n}\right).$$

Логарифмируя стоящее в конце правой части формулы произведение, получаем

$$\ln\left[1\cdot\left(1-\frac{1}{n}\right)\cdot\left(1-\frac{2}{n}\right)\cdot\ldots\cdot\left(1-\frac{k-1}{n}\right)\right]=\ln\left(1-\frac{1}{n}\right)+\ln\left(1-\frac{2}{n}\right)+\ldots+\ln\left(1-\frac{k-1}{n}\right)=$$

$$=\left[-\frac{1}{n}+O\left(\frac{1}{n}\right)^2\right]+\left[-\frac{2}{n}+O\left(\frac{2}{n}\right)^2\right]+\ldots+\left[-\frac{k-1}{n}+O\left(\frac{k-1}{n}\right)^2\right]=$$

$$=\left[-\frac{1}{n}-\frac{2}{n}-\ldots-\frac{k-1}{n}\right]+k\cdot O\left(\frac{k-1}{n}\right)^2=-\frac{k(k-1)}{n}+k\cdot O\left(\frac{k}{n}\right)^2=-\frac{k^2}{2n}+O\left(\frac{k^3}{n^2}\right),$$

откуда и следует (20).

Пусть теперь k растёт с ростом n линейно, так что $k/n\to\rho$, где $0<\rho<1$. Тогда для C_n^k имеет место следующая асимптотика

$$C_n^k\sim\frac{1}{\sqrt{2\pi n\rho(1-\rho)}}2^{-n(\rho\log_2\rho+(1-\rho)\log_2(1-\rho))},\ n\to\infty,\ k/n\to\rho. \quad (21)$$

Для доказательства (21) достаточно применить формулу Стирлинга

$$C_n^k=\frac{n!}{k!(n-k)!}=\frac{n!}{(\rho n)!\big((1-\rho)n\big)!}\sim$$

$$\sim\frac{\sqrt{2\pi n}\,n^n e^{-n}}{\sqrt{2\pi\rho n}\,(\rho n)^{\rho n}\sqrt{2\pi(1-\rho)n}\,\big((1-\rho)n\big)^{(1-\rho)n}}=$$

$$= \frac{1}{\sqrt{2n\rho(1-\rho)}}\left(\frac{1}{\rho^{\rho}(1-\rho)^{1-\rho}}\right)^{n} = \frac{e^{-n(\rho\ln\rho+(1-\rho)\ln(1-\rho))}}{\sqrt{2\pi n\rho(1\text{-}\rho)}} =$$

$$= \frac{2^{-n(\rho\log_2\rho+(1-\rho)\log_2(1-\rho))}}{\sqrt{2\pi n\rho(1-\rho)}}.$$

В комбинаторных задачах нередко используется следующее вытекающее из (21) асимптотическое равенство

$$C_n^{\rho n} = 2^{-n(\rho\log_2\rho+(1-\rho)\log_2(1-\rho)+o(1))} = 2^{n\left(H_2(\rho)+o(1)\right)},\ n\to\infty, \qquad (22)$$

где $H_2(\rho) = -\rho\log_2\rho - (1-\rho)\log_2(1-\rho)$ — функция, особенно часто возникающая в теории информации и называемая двоичной энтропией. Её график представлен на рис. 5.1.1.

Оценим теперь часто встречающиеся в комбинаторном анализе суммы вида $\sum_{i=0}^{k} C_n^i$. При $k < n/2$ для них справедливы следующие оценки

$$C_n^k \le \sum_{i=0}^{k} C_n^i < \frac{n-k}{n-2k} C_n^k. \qquad (23)$$

Нижняя оценка тривиальна. Для получения верхней оценки воспользуемся формулой для суммы геометрической прогрессии. При $i \le k < n/2$ имеем

$$\frac{C_n^{i-1}}{C_n^i} = \frac{i}{n-i+1} < \frac{i}{n-i} < \frac{k}{n-k} < 1.$$

Отсюда получаем

$$\sum_{i=0}^{k} C_n^i < C_n^k\left(1+\frac{k}{n-k}+\left(\frac{k}{n-k}\right)^2+\ldots+\left(\frac{k}{n-k}\right)^k\right) = C_n^k\frac{1-\left(\frac{k}{n-k}\right)^{k+1}}{1-\frac{k}{n-k}} <$$

$$< C_n^k\frac{1}{1-\frac{k}{n-k}} = \frac{n-k}{n-2k}C_n^k.$$

Оценка (23) показывает, в частности, что при $n\to\infty$ и $k = o(n)$ величина суммы асимптотически совпадает со своим старшим членом. Если же $k = \rho n$, то (22) и (23) дают

$$\sum_{i=0}^{\rho n} C_n^i = 2^{-n(\rho\log_2\rho+(1-\rho)\log_2(1-\rho)+o(1))} = 2^{n\left(H_2(\rho)+o(1)\right)},\ n\to\infty. \qquad (24)$$

Неравенства. В заключение приведём для справок ряд неравенств, используемых в комбинаторном анализе при получении различных оценок. Все они вытекают из выпуклости соответствующих функций, кроме того, (26) и (27) сразу следуют из (25).

$$1+x \le e^x, \text{ где } x \in R; \qquad 1-x \le e^{-x}, \text{ где } x \in R; \tag{25}$$

$$\ln(1+x) \le x, \text{ где } x \ge -1; \qquad \ln x \le x-1, \text{ где } x > 0; \tag{26}$$

$$(1+x)^r \le e^{rx}, \text{ где } x \ge -1, r \ge 0; \qquad (1-x)^r \le e^{-rx}, \text{ где } x \le 1, r \ge 0; \tag{27}$$

$$(1+x)^r \ge 1+rx, \text{ где } x \ge -1, r \ge 1; \qquad (1-x)^r \ge 1-rx, \text{ где } x \le 1, r \ge 1. \tag{28}$$

Часто используется также следующий известный из математического анализа предел. Если $\alpha \to 0$, $\beta \to \infty$ так, что $\alpha \cdot \beta = \lambda$, где λ — постоянная величина, то

$$(1-\alpha)^\beta \to e^{-\lambda}. \tag{29}$$

Вопросы для самопроверки

1. Какое из трёх соотношений имеет место при $n \to \infty$?

   а) $C_n^{[n/2]} = O(2^n/n^2)$; б) $C_n^{[n/2]} = O(2^n/n)$; в) $C_n^{[n/2]} = O(2^n/\sqrt{n})$.

2. Какое из трёх соотношений имеет место при $n \to \infty$?

   а) $\sum\limits_{i=-\sqrt{n}}^{\sqrt{n}} C_n^{[n/2]+i} = o(2^n)$; б) $\sum\limits_{i=-\sqrt{n}}^{\sqrt{n}} C_n^{[n/2]+i} = O(2^n)$;

   в) $\sum\limits_{i=-\sqrt{n}}^{\sqrt{n}} C_n^{[n/2]+i} \sim 2^n$.

Ответы: 1 — в, 2 — б.

Задачи для самостоятельного решения

1. В купе железнодорожного вагона имеются два противоположных дивана, каждый с 5 пронумерованными местами. Из 10 пассажиров этого купе четверо желают сидеть лицом по ходу движения, двое — спиной, а остальным безразлично как сидеть. Сколько способов размещения пассажиров с учётом их желаний существует?
2. Сколько существует различных начальных ситуаций при игре в «подкидного дурака», где каждый из четырёх играющих получает при сдаче по 6 карт, а оставшиеся 12 карт лежат в определённом порядке?
3. Восемь молодых людей, девушек и юношей, хотят разделиться на две команды по четыре человека для игры в городки. Сколькими способами они могут это сделать, если

   а) девушек пятеро, а юношей трое, и в каждой команде должен быть хотя бы один юноша?

   б) девушек четверо и юношей четверо, а в каждой команде должно быть по две девушки и по два юноши?

4. Чему равно число
   а) всех бинарных $(m \times n)$-матриц?
   б) бинарных $(m \times n)$-матриц с k единицами и $mn-k$ нулями?
   в) бинарных $(m \times n)$-матриц, не имеющих одинаковых строк?
   г) бинарных $(m \times n)$-матриц, не имеющих нулевых строк?
   д) бинарных $(m \times n)$-матриц без нулевых и одинаковых строк.
5. Сколькими способами можно n шаров разместить по m урнам так, что чтобы в m_1 урнах находилось по n_1 шаров, в m_2 урнах находилось по n_2 шаров и т. д., в m_k урнах находилось по n_k шаров ($m_1 + m_2 + \ldots + m_k = m$, $m_1 n_1 + m_2 n_2 + \ldots + m_k n_k = n$), если
   а) шары различимы и урны различимы;
   б) шары различимы, а урны неразличимы;
   в) шары неразличимы, а урны различимы;
   г) и шары, и урны неразличимы.
6. На окружности выбрано n точек общего положения, и каждая пара соединена хордой. Сколько точек пересечения хорд внутри окружности при этом возникло? (Общее положение означает, что через каждую точку пересечения хорд внутри окружности проходит ровно две хорды. Такое расположение с вероятностью единица возникает при случайном выборе точек на окружности).
7. Чему равно число 4-значных десятичных чисел, у которых
   а) каждая последующая цифра меньше предыдущей?
   б) каждая последующая цифра не больше предыдущей?
8. Сколько существует бинарных наборов длины $(n+m)$ с n единицами и m нулями, в которых никакие 2 единицы не стоят рядом? ($n \le m+1$)
9. Сколькими способами n человек можно рассадить за круглым столом с $n+m$ пронумерованными местами так, чтобы никакие двое не сидели рядом? ($m \ge n$)
10. Сколько существует возможных вариантов выпадения при одновременном бросании пяти одинаковых игральных костей?
11. Сколькими способами 2 монеты по 10 рублей, 4 монеты по 5 рублей и 5 монет по 1 рублю могут быть разложены по 3 разным кошелькам? (Пустые кошельки допускаются)
12. Сколько решений в целых неотрицательных числах имеет неравенство $x_1 + x_2 + \ldots + x_n \le m$?
13. Функцией Эйлера $\varphi(n)$ для натуральных n называется число натуральных чисел, не превосходящих n и взаимно простых с n. Показать, что при $n > 1$ $\varphi(n) = n \prod_{i=1}^{k} (1 - \frac{1}{p_i})$, где $p_1, p_2, \ldots, p_k$ — все различные простые делители числа n.

14. Доказать следующие тождества:

а) $\sum_{i=m}^{n} C_i^m = C_{n+1}^{m+1}$; б) $\sum_{k=0}^{n}\sum_{r=0}^{k} C_n^k C_k^r = 3^n$.

15. Для всех натуральных n доказать справедливость двойного неравенства

$$\frac{4^n}{2\sqrt{n}} \le C_{2n}^n \le \frac{4^n}{\sqrt{3n+1}}.$$

16. Мэри мечтает выйти замуж за умного, красивого и богатого мужчину. Она считает, что каждый из 20 её поклонников обладает хотя бы одним из этих качеств, причём среди них 11 богатых, 10 умных, 9 красивых, 3 умных и богатых, 4 красивых и умных, 4 богатых и красивых. Есть ли среди поклонников мужчина её мечты?
17. Сколько натуральных чисел, не превосходящих 1000, не делятся ни на одно из чисел 6,10,15?
18. Компания из n мужчин заходит в бар, где каждый оставляет в гардеробе зонт и шляпу. Выходя из бара, компания разбирает зонты и шляпы случайным образом. Найти вероятность того, что

    а) ровно l мужчин окажутся со своим зонтом и своей шляпой;

    б) ровно у l мужчин окажется хотя бы одна из их вещей.

19. За круглым столом требуется рассадить n супружеских пар в соответствии с правилами этикета, согласно которым мужчины и женщины должны чередоваться и никакие двое супругов не должны сидеть рядом. Сколькими способами это может быть осуществлено, если расположения, получаемые круговым сдвигом всей компании, считать одинаковыми? (Эта задача известна среди математиков с конца девятнадцатого века под названием «Le probléme des ménages». Успех в её решении связан с использованием формулы включения и исключения).
20. Сколькими способами можно раздать 12 одинаковых монет пяти нищим так, чтобы каждый получил не менее одной, но не более 3 монет?
21. Номера билетов состоят из 6 десятичных цифр, так что всего имеется 10^6 номеров. Найти число «счастливых» номеров, у которых сумма первых трёх цифр равна сумме последних трёх цифр.
22. Найти вероятность того, что две наугад взятые костяшки домино можно приставить друг к другу.
23. Из перетасованной колоды (36 карт) извлекаются 3 карты. Какова вероятность, что сумма их очков будет равна 21 (валет — 2, дама — 3, король — 4, туз — 11, остальные 6, 7, 8, 9, 10)?
24. Используется колода из 52 карт (достоинства карт начинаются с «двойки»). Карточные комбинации в вопросах задачи взяты из игры в покер, в которой играющий получает при сдаче 5 карт. Найти вероятность получить:

    а) пять карт одной масти (флеш);

    б) четыре карты из пяти одинакового достоинства (каре);

в) три карты одного достоинства и две другого (фул);
г) пять последовательных карт не одной масти (стрит);
д) три карты одного достоинства и две карты других различных между собой достоинств (тройка);
е) две карты одного достоинства и две карты другого достоинства плюс карта отличного от них достоинства (две двойки);
ж) две карты одного достоинства и три карты других различных между собой достоинств (двойка).

25. Сколькими способами можно подняться по лестнице с n ступеньками, преодолевая каждым шагом одну или две ступеньки?
26. Найти формулу n-го члена последовательности, заданной рекуррентно: $a_{n+1} = a_n + n + 1$, $a_0 = 1$.
27. Найти формулу n-го члена последовательности, заданной рекуррентно: $u_{n+1} = u_n + 2^n$, $u_0 = 0$.
28. Найти a_n и b_n из системы рекуррентных соотношений $\begin{cases} a_{n+1} = 3a_n + b_n \\ b_{n+1} = -a_n + b_n \end{cases}$ с заданными начальными условиями $a_0 = 14$, $b_0 = -6$.
29. С помощью производящих функций найти формулу для суммы кубов первых n натуральных чисел.
30. Задача о баллотировке (У. Уитворт, 1878, Ж. Бертран, 1887). Пусть на выборах за кандидата А подано m голосов, а за кандидата В — n голосов, причём $m \geq n$. Пусть, далее при подсчёте голосов $m+n$ бюллетеней вынимаются из урны случайным образом. Какова вероятность, что в процессе подсчёта голосов кандидат А ни разу не уступит кандидату В?
31. Сколькими способами $2n$ человек различного роста могут быть построены в две шеренги так, чтобы в каждой шеренге они стояли по росту, и каждый стоящий в первой шеренге был выше стоящего за ним?
32. Доказать асимптотическое равенство

$$C_{n+k}^m \Big/ C_n^m \sim e^{mk/n}, \quad n \to \infty, \quad k = o(n), \quad m = o(n).$$

Литература

1. *Гаврилов Г. П., Сапоженко А. А.* Задачи и упражнения по курсу дискретной математики. М.: Наука, 2004.
2. *Гульден Я., Джексон Д.* Перечислительная комбинаторика. М.: Наука, 1990.
3. *Грэхем Р., Кнут Д., Паташник О.* Конкретная математика. М: Мир, 1998.
4. *Егорычев Г. П.* Интегральное представление и вычисление комбинаторных сумм. Новосибирск: Наука, 1977.
5. *Журавлёв Ю. И., Флёров Ю. А., Федько О. С., Дадашев Т. М.* Сборник задач по дискретному анализу. М.: МФТИ, 2004.
6. *Кнут Д. Э.* Искусство программирования. Т. 1. М.: Вильямс, 2000.
7. Комбинаторный анализ. Задачи и упражнения (под ред. К. А. Рыбникова). М.: Наука, 1982.

8. *Леонтьев В. К.* Избранные задачи комбинаторного анализа. М: Изд-во МГТУ им. Н. Э. Баумана, 2001.
9. *Липский В.* Комбинаторика для программистов. М.: Мир, 1988.
10. Перечислительные задачи комбинаторного анализа (сб. переводов под ред. Г. П. Гаврилова). М.: Наука, 1979.
11. *Платонов М. Л.* Комбинаторные числа класса отображений и их приложения. М.: Наука, 1979.
12. Прикладная комбинаторная математика (сб. статей под ред. Э. Беккенбаха). М.: Мир, 1968.
13. *Райзер Г. Дж.* Комбинаторная математика. М: Мир, 1966.
14. *Риордан Дж.* Введение в комбинаторный анализ. М: ИЛ, 1963.
15. *Риордан Дж.* Комбинаторные тождества. М.: Наука, 1982.
16. *Рыбников К. А.* Введение в комбинаторный анализ. М.: Изд-во Моск. ун-та, 1985.
17. *Сачков В. Н.* Введение в комбинаторные методы дискретной математики. М.: МЦНМО, 2004.
18. *Стенли Р.* Перечислительная комбинаторика. Т. 1. М.: Мир, 1990.
19. *Стенли Р.* Перечислительная комбинаторика. Т. 2. М.: Мир, 2005.
20. *Феллер В.* Введение в теорию вероятностей и её приложения. Т. 1. 2-е изд. М.: Книжный дом «Либроком»/URSS, 2010.
21. *Эйлер Л.* Введение в анализ бесконечно малых. Т. 1. М.: ОНТИ, 1990.
22. *Эндрюс Г.* Теория разбиений. М.: Наука, 1982.
23. *Lovász L.* Combinatorial Problems and Exercises. Budapest: Akadémiai Kiadó, 1993.

Комментарии к литературе

Монография [14] в течение долгого времени была основным русскоязычным источником информации по теории перечисления. В определённой мере она, а также более поздняя книга [15] сохраняют своё значение, но в качестве начального курса лучше использовать, например, [13]. Для дальнейшего более глубокого изучения предмета может быть рекомендованы монография [3]. Прекрасным вводным курсом может служить [6], а также [9]. Книги [18] и [19] являются достаточно полными курсами современной алгебраической комбинаторики. Связи перечислительной комбинаторики с теорией вероятностей отмечены в [20]. Книга [16] с задачником [7] является солидным учебником университетского типа, а [17] — продвинутым учебником. Большое число задач перечислительного характера с решениями можно найти в [1]. Задачник [6] составлен на основе опыта занятий со студентами МФТИ. Сборник [12] содержит важную работу де Брейна, посвящённую теории перечисления Пойа. С оригинальными работами других классиков комбинаторного анализа, включая и работу самого Пойа, можно ознакомиться по сборнику [10]. Монография [22] посвящена подсчёту числа разбиений натуральных чисел на суммы — задаче, к решению которой метод производящих функций впервые был применён Эйлером [21]. Книга [2] посвящена рассмотрению метода производящих функций в самом общем виде, и она не для лёгкого чтения. Изучение монографии [4] требует знакомства с аналитическими функциями. Значительное число задач, использующих развитые в [4] методы, можно найти в [8]. Сборник задач по комбинаторному анализу [23] получил высокую оценку профессионалов во всём мире.

Глава 2

Булевы функции

2.1. Булевы функции и логические связки

Если каждому элементу множества X поставлен в соответствие единственный элемент множества Y, то говорят, что задано отображение из X в Y или функция $f: X \to Y$. Общепринятой записью функции является также обозначение $y = f(x)$, где $x \in X$, $y \in Y$. В математическом анализе в качестве множеств X и Y выступает, как правило, множество действительных чисел или его интервалы. В дискретной математике часто рассматриваются функции, область определения и множество значений которых являются конечными множествами. Особенно важными являются *булевы функции*, отображающие некоторую декартову степень множества $\{0,1\}$ во множество $\{0,1\}$. Они названы так в честь Джорджа Буля (1815–1864) — английского математика-самоучки, заложившего основы символического исчисления, которое теперь называется *булевой алгеброй*.

> Это исчисление появилось в результате попыток исследовать фундаментальные законы тех операций, которые совершает разум в процессе рассуждений; записать их в символическом языке Исчисления и на этой основе создать науку Логики и построить её метод; использовать сам этот метод в качестве исходного пункта для развития общего метода применения математической теории к исследованию Вероятностей...
>
> *Джордж Буль «Исследование законов мышления» (1854)*

Судьба созданного Булем исчисления оказалась воистину блестящей. Оно нашло применение в теории множеств, математической логике, теории вероятностей и кибернетике. Исчисление имеет три основные операции, две бинарных и одну унарную. Эти операции носят абстрактный характер, постулируются лишь их свойства. В теории множеств этими операциями становятся объединение, пересечение и дополнение. В математической логике в качестве этих операций выступают *дизъюнкция*, *конъюнкция* и *отрицание*. К изучению этих операций над булевыми функциями мы теперь и переходим. Из-за той роли, которую булевы функции играют в математической логике, их часто называют также *функциями алгебры логики*.

Булевой функцией от n переменных $f: \{0,1\}^n \to \{0,1\}$ называется функция $f(x_1,\ldots,x_n)$, определённая на множестве всех двоичных наборов длины n и принимающая на каждом из них значение 0 или 1.

Наборы, на которых булева функция принимает единичное и нулевое значение, называют соответственно её *единичными* и *нулевыми наборами*. Множество нулевых наборов функции f часто обозначают как $f^{-1}(0)$, а единичных — как $f^{-1}(0)$. Для множества единичных наборов будем также использовать обозначение N_f. Так как имеется 2^n двоичных наборов длины n, на каждом из которых булева функция принимает одно из двух значений, число булевых функций от n переменных равно 2^{2^n}. Множество булевых функций от n переменных обозначают как P_2^n, а всё множество булевых функций — как P_2, $P_2 = \bigcup_{n=0}^{\infty} P_2^n$. При этом под функциями, зависящими от нулевого числа переменных, понимаются константы 0 и 1. Множество P_2 счётно.

Заметим также, что рассматривая двоичный набор как характеристический вектор, булеву функцию можно считать заданной на подмножествах n-элементного множества. При этом единичными значениями функции выделяется некоторое семейство подмножеств.

Булева функция $f: \{0, 1\}^n \to \{0, 1\}$ может быть задана таблично, т. е. для каждого из 2^n наборов значений переменных, выписываемых обычно в лексикографическом порядке (по возрастанию двоичных чисел), задаётся значение функции на этом наборе. Такую таблицу принято называть *таблицей истинности*.

При $n = 1$ имеем 4 булевы функции. Все они представлены в табл. 1.

x \ $f(x)$	**0**	**1**	x	$\bar{x}$
0	0	1	0	1
1	0	1	1	0

Таблица 1

При этом функция x называется *тождественной*, а функция $\bar{x}$ — *отрицанием* x (читается «не x»). В математической логике для отрицания используется также обозначение $\neg x$. Заметим, что $\bar{\bar{x}} = x$.

Булевы функции от одной переменной 0 и 1 в действительности являются константами и не зависят от переменной x. Переменная x является в данном случае *фиктивной* или, как говорят, *несущественной* переменной. В общем случае переменная x_i функции $f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)$ назы-

вается *существенной*, если существует такой набор значений остальных переменных $(\alpha_1,\ldots,\alpha_{i-1},\alpha_{i+1},\ldots,\alpha_n)\in\{0,1\}^{n-1}$, что

$$f(\alpha_1,\ldots,\alpha_{i-1},0,\alpha_{i+1},\ldots,\alpha_n)\neq f(\alpha_1,\ldots,\alpha_{i-1},1,\alpha_{i+1},\ldots,\alpha_n).$$

В противном случае она называется *фиктивной*, или, как говорят, *несущественной* переменной и может быть опущена, так как не влияет на значение функции. Заметим, что в P_2^n включаются и функции, фактически зависящие от меньшего числа переменных, т. е. не все переменные которых существенны. Отождествив между собой функции, получающиеся одна из другой изъятием или добавлением несущественных переменных, имеем $P_2^i\subseteq P_2^j$, при $i\le j$.

Булевых функций от двух переменных имеется уже 16. При этом наибольшее значение и специальные названия имеют семь из них (табл. 2):

x_1	x_2 \ $f(x_1,x_2)$	$x_1 \,\&\, x_2$	$x_1 \vee x_2$	$x_1 \to x_2$	$x_1 \sim x_2$	$x_1 \oplus x_2$	$x_1 \mid x_2$	$x_1 \downarrow x_2$
0	0	0	0	1	1	0	1	1
0	1	0	1	1	0	1	1	0
1	0	0	1	0	0	1	1	0
1	1	1	1	1	1	0	0	0

Таблица 2

$x_1 \,\&\, x_2$ – *конъюнкция* (связка «и», логическое умножение), обозначается также $x_1 \wedge x_2$, $x_1 \cdot x_2$ или просто x_1x_2;

$x_1 \vee x_2$ – *дизъюнкция* (включающее «или», логическое сложение);

$x_1 \to x_2$ – *импликация* (логическое следование);

$x_1 \sim x_2$ – *эквивалентность*;

$x_1 \oplus x_2$ – *сложение по модулю 2* (разделительное «или»);

$x_1 \mid x_2$ – *штрих Шеффера (не «и»)*;

$x_1 \downarrow x_2$ – *стрелка Пирса (не «или»)*.

Булевы функции возникли в результате смыслового анализа конструкций естественных языков. Основой естественных языков являются *высказывания*, т. е. некоторые утверждения, о которых есть смысл говорить, что они истинны или ложны. Например, высказывание «Я живу в Москве» истинно, если Вы действительно проживаете в Москве, и ложно в противном случае. В естественном языке подобные простые высказывания соединяются в сложные при помощи частицы «не», союзов и языковых конструкций «и», «или», «если ..., то», «в том и только в том случае, если»,

«либо ..., либо» и их эквивалентов. Например, высказывание «Я живу в Москве и учусь в Московском государственном университете» — сложное высказывание, составленное из двух простых.

Если каждому простому высказыванию поставить в соответствие булеву переменную, договорившись, что значению высказывания «истина» соответствует её единичное значение, а значению высказывания «ложь» — нулевое, то появляется возможность выражать сложные высказывания с помощью булевых операций. Например, поставив в соответствие простому высказыванию «Я живу в Москве» булеву переменную A, а простому высказыванию «Я учусь в Московском государственном университете» — булеву переменную B, приведённое выше сложное высказывание можно записать как $A \& B$. Сложное же высказывание «Я живу в Москве, но не учусь в Московском государственном университете» запишется как $A \& \overline{B}$. Подобный формализм называется *исчислением высказываний*.

Отрицание, конъюнкция, дизъюнкция, импликация, эквивалентность и сложение по модулю 2 используются в качестве логических связок в исчислении высказываний, являясь эквивалентами соответственно частицы «не», союзов и языковых конструкций «и», «или», «если ..., то», «в том и только в том случае, если», «либо ..., либо» естественного языка. Соответствие булевых операций логическим связкам естественного языка приведено в табл. 3.

$\neg P$	$P \& Q$	$P \vee Q$	$P \to Q$	$P \sim Q$	$P \oplus Q$
не P	Q и Q	P или Q	если P, то Q	P в том и только в том случае, если Q	либо P, либо Q

Таблица 3

Рассмотрим использование булевых операций в качестве логических связок подробнее на другом примере. Пусть P обозначает высказывание «Я пойду гулять», а Q — «Я буду заниматься математикой». Тогда следующие булевы выражения будут обозначать высказывания:

$\overline{P}$ — «Я не пойду гулять»;

$\overline{Q}$ — «Я не буду заниматься математикой»;

$P \vee Q$ — «Я пойду гулять или буду заниматься математикой» (причём допускается, что я буду делать и то, и другое, хотя в естественном языке союз «или» может использоваться как в разделительном, так и в неразделительном смысле, что оставляет некоторую неопределённость);

$P \oplus Q$ — «Я либо пойду гулять, либо буду заниматься математикой» (здесь возможность и того, и другого уже исключается);

$P \& Q$ — «Я пойду гулять и буду заниматься математикой» (очевидно, на прогулке);

$\bar{P} \to Q$ — «Если я не пойду гулять, то буду заниматься математикой» (если в естественном языке подобная фраза часто подразумевает, что прогулка исключает занятие математикой, то написанная формула в соответствии с таблицей истинности допускает занятие математикой также и в случае прогулки);

$\bar{P} \sim Q$ — «Я буду заниматься математикой в том и только в том случае, если я не пойду гулять».

Приведённые примеры наглядно демонстрируют смысловую точность исчисления высказываний на фоне двусмысленности естественного языка. Здесь стоит заметить, что, в отличие от большинства современных европейских языков, латинский язык, долгое время служивший языком науки, имел союз «vel» для включающего «или» и союз «aut» для разделительного «или». От первого из этих союзов и был образован символ « $\vee$ » для обозначения дизъюнкции.

В то же время естественный язык намного богаче эмоционально-смысловыми оттенками. Выражение « A , но и B » отличается эмоциональной окраской от выражения « A и B », но в исчислении высказываний оба выражения запишутся одинаково: $A \& B$ или $B \& A$, что в исчислении высказываний эквивалентно. В естественном языке, однако, порядок простых высказываний, входящих в состав сложного высказывания, может выражать временную и, как результат, возможную причинно-следственную связь явлений. Например, высказывание «Один мотор заглох, и самолёт начал терять высоту», рассматриваемое как $A \& B$, при замене на $B \& A$ теряет смысл, хотя в исчислении высказываний два выражения эквивалентны.

Подобных проблем не возникает при записи с помощью логических связок математических утверждений, так как их истинность не зависит от времени. Поэтому булевы функции и нашли себе применение в математической логике. Но если в математической логике булевы функции используются в качестве аппарата, то в дискретной математике они сами являются предметом изучения и притом с двух точек зрения, комбинаторной и функциональной. В первом случае изучается взаимное расположение нулевых и единичных наборов функции, а во втором — порождение новых функций путём суперпозиций из заданного множества функций.

Чтобы понять, как булевы функции используются в математической логике, рассмотрим взятую из [13] задачу, составленную в духе «криминально-детективного жанра», в которой требуется проверить логическую правильность сделанного заключения.

«Если Джонс не встречал этой ночью Смита, то либо Смит был убийцей, либо Джонс лжёт. Если Смит не был убийцей, то Джонс не встречал этой ночью Смита, и убийство имело место после полуночи. Если убийство имело место после полуночи, то либо Смит был убийцей, либо Джонс лжёт. Следовательно, Смит был убийцей».

Считая все три приведённые посылки безусловно верными, требуется проверить, вытекает ли из них сделанное заключение. Для решения этой задачи введём буквенные обозначения для встречающихся здесь элементарных высказываний:

P — Джонс не встречал этой ночью Смита;
Q — Смит был убийцей;
R — Джонс лжёт;
S — убийство имело место после полуночи.

Тогда условия задачи запишутся в виде

1) $P \to (Q \oplus R)$;
2) $\bar{Q} \to (P \,\&\, S)$;
3) $S \to (Q \oplus R)$.

$$\frac{}{\overset{?}{Q = 1}}$$

Для проверки правильности логического вывода, нужно выяснить, существует ли такое приписывание значений используемым в нём логическим переменным, при котором все посылки будут истинны, а заключение ложно. Если такой набор истинностных значений переменных существует, то это, означает логическую неправомерность заключения. Если же отрицание заключения несовместимо с истинностью посылок, то заключение является логическим следствием посылок.

Чтобы сократить перебор, поступим следующим образом. Положим, что заключение ложно, и посмотрим, какие значения могут принимать остальные переменные при условии истинности посылок. Пусть Q ложно. Тогда из второй посылки следует, что P и S истинны, а для выполнения первой и третьей посылок достаточно положить, что R истинно. Таким образом, заключение логически не вытекает из посылок.

Теперь слегка изменим условия задачи. Пусть первые две посылки останутся неизменными, а третья будет такой «Если убийство имело место после полуночи, то либо Смит не был убийцей, либо Джонс лжёт», что в наших обозначениях запишется как $S \to (\bar{Q} \oplus R)$. Теперь, если Q ложно, то из второй посылки, по-прежнему, следует, что P и S истинны, но из первой посылки следует, что R истинно, а из третьей — что R ложно. Полученное противоречие показывает, что отрицание заключения несовместимо с истинностью посылок, т. е. вывод о том, что Смит убийца является в этом случае логически правомерным.

Позволим себе здесь на минуту отвлечься от математики и привести любопытную цитату из произведения основателя дстективного жанра американского писателя Эдгара По (1809–1849), который устами своего героя,

мастера логического распутывания детективных головоломок Огюста Дюпена заявляет:

> Как математик и поэт он должен рассуждать хорошо; будь он только математиком, он не умел бы рассуждать вовсе .... Я подвергаю сомнению годность, а следовательно, и ценность того разума, который культивируется в любом специфическом виде, кроме абстрактно логического. Особому сомнению я подвергаю разум, взращенный на математических штудиях. Математика — это наука о форме и количестве, математический ум — это всего лишь логика в приложении к наблюдениям над формой и количеством.
>
> *Эдгар Аллан По «Похищенное письмо» (1845)*

Тезис о неспособности математика к абстрактным логическим рассуждениям вряд ли нуждается в опровержении. Гениальному поэту и новеллисту нельзя, однако, отказать в прозорливости. Он верно почувствовал отсутствие в современной ему математике дисциплины, которая бы изучала абстрактно логические отношения. И действительно, уже через два года после появления процитированных строк вышли первые работы Дж. Буля и де Моргана, положившие начало развитию математической логики.

Предикаты и кванторы. Чтобы иметь возможность компактно записывать математические утверждения на языке математической логики, помимо высказываний необходимы *предикаты* и *кванторы*. Функция $P(x_1,\ldots,x_n)$, переменные которой принимают значения из некоторого множества M, называется *n-местным предикатом*, если при подстановке вместо символов переменных конкретных значений из множества M она принимает значение 1 («истина») или 0 («ложь»), т. е. становится высказыванием. Множество тех наборов значений аргументов предиката, на которых он принимает значение 1, называется *областью истинности предиката*.

Пусть, например, $P(x)$ обозначает одноместный предикат «x — простое число», определённый на множестве натуральных чисел. Тогда $P(2)=1$, $P(3)=1$, $P(4)=0$. Область истинности этого предиката — множество простых чисел.

Пусть, теперь, $P(x,y)$ — двухместный предикат «$x \le y$», где x и y — действительные числа. Тогда $P(3,5)=1$, $P(5,3)=0$. Область истинности любого двухместного предиката представляет собой бинарное отношение.

Типичное математическое утверждение заключается либо в утверждении существования в некотором множестве элемента с заданным свойством, либо в утверждении, что этим свойством обладают все элементы множества. Для записи подобных утверждений используются *квантор существования* $\exists x$ и *квантор всеобщности* $\forall x$.

Запись $\exists x P(x)$ означает, что область истинности предиката $P(x)$ не пуста, а $\forall x P(x)$ — что область истинности предиката $P(x)$ совпадает с

областью его определения. Выражение $\exists x P(x)$ читается как "существует такое x, что $P(x)$", а выражение $\forall x P(x)$ — «для всех x, $P(x)$».

В качестве примера запишем на языке математической логики приведённое во Вводной главе определение алгебраической группы. Условия ассоциативности, существования единичного и обратного элементов могут быть теперь записаны как

$$\forall a \forall b \forall c \ (ab)c = a(bc);$$

$$\exists e \forall a \qquad ea = ae = a;$$

$$\forall a \exists a^{-1} \quad a^{-1}a = aa^{-1} = e.$$

Вопросы для самопроверки

1. Сколько существует булевых функций от 3 переменных?
   а) 64; б) 256; в) 324.
2. Пусть P обозначает высказывание «будет дождь», а Q — «я пойду гулять». Как будет выглядеть в этих обозначениях высказывание «если не будет дождя, то я пойду гулять»?
   а) $Q \to P$; б) $Q \sim P$; в) $\overline{P} \to Q$.

Ответы: 1 — б, 2 — в.

2.2. Формулы и преобразования

Так как область изменения каждой переменной булевой функции есть множество $\{0,1\}$, а множество её значений также принадлежит этому множеству, то это позволяет вместо каждой из переменных некоторой булевой функции подставлять другие функции, получая, таким образом, из имеющегося запаса функций новые функции и выражая одни функции через другие. Такая подстановка в математике называется *суперпозицией* функций.

Примерами такой суперпозиции является отмеченное в определении задание штриха Шеффера и стрелки Пирса как отрицание соответственно конъюнкции и дизъюнкции:

$$x_1 \mid x_2 = \overline{x_1 x_2};$$
$$x_1 \downarrow x_2 = \overline{x_1 \vee x_2}.$$

В первом случае вычисляется конъюнкция переменных x_1 и x_2 и к результату применяется отрицание, т. е. имеет место суперпозиция конъюнкции и отрицания, а во втором используется суперпозиция дизъюнкции и отрицания.

В теории булевых функций операция суперпозиции важна ещё и потому, что при практической реализации булевых функций схемами суперпозиция осуществляется, когда выход одного из элементов схемы подаётся

на вход другого. Последовательное использование операции суперпозиции, приводит к аналитическому заданию булевых функций *формулами.*

Пусть $A = \{f_1, f_2, \ldots\}$ — произвольная система булевых функций $f_i \in P_2$, конечная или бесконечная. Тогда множество *булевых формул над* A индуктивно может быть определено следующим образом:

1) *каждая переменная* x_i *является формулой*;
2) *если* $f(x_1,\ldots,x_k) \in A$ *и* $F_1,\ldots,F_k$ *суть формулы (не обязательно различные), то* $f(F_1,\ldots,F_k)$ *есть формула*;
3) *других формул нет.*

Штрих Шеффера и стрелка Пирса выражаются, таким образом, формулами над $A = \{\vee, \&, \neg\}$. Сложение по модулю 2, импликация и эквивалентность также могут быть выражены формулами с помощью дизъюнкции, конъюнкции и отрицания. При этом согласно общепринятому соглашению при записи формул *первыми выполняются отрицания переменных, а конъюнкция выполняется раньше дизъюнкции и сложения по модулю 2.* Это позволяет опускать часть скобок при записи формул:

$$x_1 \oplus x_2 = (x_1\bar{x}_2) \vee (\bar{x}_1 x_2) = x_1\bar{x}_2 \vee \bar{x}_1 x_2;$$
$$x_1 \to x_2 = \bar{x}_1 \vee x_2;$$
$$x_1 \sim x_2 = (x_1 x_2) \vee (\bar{x}_1\bar{x}_2) = x_1 x_2 \vee \bar{x}_1\bar{x}_2.,$$

Справедливость этих формул может быть проверена с помощью таблиц истинности непосредственным перебором всех возможных наборов значений x_1 и x_2.

Заметим также что, эквивалентность может быть выражена через импликацию и конъюнкцию как

$$x_1 \sim x_2 = (x_1 \to x_2) \& (x_2 \to x_1),$$

что вполне отвечает привычному представлению о том, что x_1 эквивалентно x_2, если x_1 влечёт x_2 и x_2 влечёт x_1.

В дальнейшем дизъюнкция, конъюнкция и отрицание станут для нас основной системой функций, через которые с помощью формул будут выражаться все остальные функции. Кроме того, нас будет интересовать также сложение по модулю 2. Поэтому рассмотрим свойства этих функций подробнее. Всюду далее A, B, C, ... — произвольные булевы формулы. Следующие тождества могут быть проверены с помощью таблиц истинности непосредственным перебором всех наборов значений входящих в них переменных:

1) правила поглощения
$$A \cdot A = A;$$
$$A \vee A = A;$$
$$A \cdot \bar{A} = 0;$$
$$A \vee \bar{A} = 1,$$

2) правила подстановки констант $A \cdot 1 = A$;
$A \vee 1 = 1$;
$A \cdot 0 = 0$;
$A \vee 0 = A$,

3) правила де Моргана

$$\overline{A \vee B} = \bar{A} \cdot \bar{B};$$
$$\overline{A \cdot B} = \bar{A} \vee \bar{B}.$$

Для дизъюнкции, конъюнкции, эквивалентности и сложения по модулю 2 выполняются также следующие алгебраические свойства.

1. *Коммутативность:*

$$A \vee B = A \vee B;$$
$$A \cdot B = A \cdot B;$$
$$A \sim B = A \sim B;$$
$$A \oplus B = A \oplus B.$$

2. *Ассоциативность:*

$$(A \vee B) \vee C = A \vee (B \vee C) = A \vee B \vee C;$$
$$(A \cdot B) \cdot C = A \cdot (B \cdot C) = A \cdot B \cdot C;$$
$$(A \oplus B) \oplus C = A \oplus (B \oplus C) = A \oplus B \oplus C.$$

Свойство ассоциативности позволяет опускать скобки в выражениях вида $A \vee B \vee C$, $A \cdot B \cdot C$, $A \oplus B \oplus C$.

3. *Дистрибутивность:*

$$A \cdot (B \vee C) = A \cdot B \vee A \cdot C$$

(конъюнкции по отношения к дизъюнкции);

$$A \vee (B \cdot C) = (A \vee B) \cdot (A \vee C)$$

(дизъюнкции по отношению к конъюнкции);

$$A \cdot (B \oplus C) = A \cdot B \oplus A \cdot C$$

(конъюнкции по отношению к сложению по модулю 2).

Преобразование булевых выражений состоит в последовательном переходе от одной формулы к другой, тождественно равной исходной. Результатом таких преобразований может быть более простая и удобная для анализа булева формула. Рассмотрим способы осуществления таких преобразований. В процессе преобразования булевых выражений, содержащих дизъюнкцию, конъюнкцию и отрицание, особенно часто используются следующие переходы:

$$A \vee AB = A \text{ (поглощение);} \quad (1)$$

$$AB \vee A\bar{B} = A \text{ (слияние),} \quad (2)$$

где A и B — произвольные булевы формулы. Слияние особенно часто используется в форме

$$xA \vee \bar{x}A = A \text{ (склеивание по переменной } x\text{).} \quad (3)$$

Полезна также формула

$$A \vee \bar{A}B = A \vee B\,. \qquad (4)$$

Доказательство этих соотношений легко следует из сформулированных выше основных свойств операций $\{\vee, \&, \neg\}$. При доказательстве двух первых соотношений используется дистрибутивность конъюнкции по отношения к дизъюнкции, а при доказательстве последнего — дистрибутивность дизъюнкции по отношению к конъюнкции:

$$A \vee AB = A(1 \vee B) = A \cdot 1 = A\,;$$

$$AB \vee A\bar{B} = A(B \vee \bar{B}) = A \cdot 1 = A\,;$$

$$A \vee \bar{A}B = (A \vee \bar{A})(A \vee B) = 1 \cdot (A \vee B) = A \vee B\,.$$

Правила де Моргана также часто используются в подобного рода преобразованиях. По индукции эти правила могут быть распространены на любое число переменных:

$$\overline{A_1 \vee A_2 \vee \ldots \vee A_n} = \bar{A}_1 \cdot \bar{A}_2 \cdot \ldots \cdot \bar{A}_n\,;$$

$$\overline{A_1 \cdot A_2 \cdot \ldots \cdot A_n} = \bar{A}_1 \vee \bar{A}_2 \vee \ldots \vee \bar{A}_n\,.$$

Отметим очевидное соответствие между булевыми операциями дизъюнкции, конъюнкции и отрицания и теоретико-множественными операциями объединения, пересечения и дополнения, проявляющееся в полном совпадении их свойств. Такое соответствие станет понятным, если каждой булевой функции от n переменных поставить во взаимно однозначное соответствие множество её единичных наборов. При этом операциям $\{\vee, \&, \neg\}$ над булевыми функциями будут соответствовать операции $\{\cup, \cap, \bar{}\}$ над соответствующими им множествами. В свете сказанного в справедливости соотношения $A \vee \bar{A}B = A \vee B$, например, можно убедиться также с помощью диаграммы Эйлера—Венна на рис. 1, где A и B являются множествами единиц соответствующих булевых выражений.

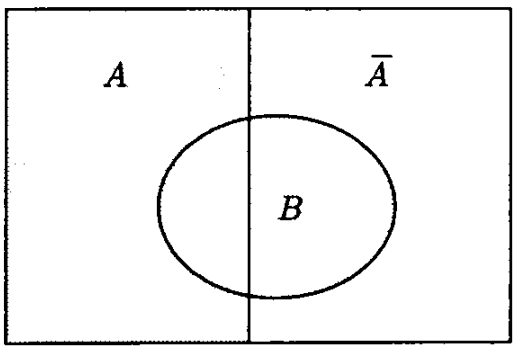


Рис. 1

Заметим также, что импликация $A \to B$ тождественно истинна в том и только в том случае, если множество единичных наборов формулы A является подмножеством единичных наборов формулы B.

Так как булевы операции $A \to B$, $A \sim B$, $A \oplus B$, $A \mid B$, $A \downarrow B$ выражаются, как уже было показано, через $\{\vee, \&, \neg\}$, то и любая формула, использующая лишь эти 8 операций, может быть преобразована в формулу, содержащую лишь $\{\vee, \&, \neg\}$.

В качестве примера рассмотрим формулу

$$(x_1 \to x_2 x_3)(x_2 \to x_1 x_3) \vee (x_1 \sim x_2).$$

Выразив импликации и эквивалентность через $\{\vee, \&, \neg\}$, имеем

$$(x_1 \to x_2 x_3)(x_2 \to x_1 x_3) \vee (x_1 \sim x_2) = (\overline{x}_1 \vee x_2 x_3)(\overline{x}_2 \vee x_1 x_3) \vee x_1 x_2 \vee \overline{x}_1 \overline{x}_2.$$

Перемножая скобки и помня, что $x \cdot \overline{x} = 0$, получаем

$$(\overline{x}_1 \vee x_2 x_3)(\overline{x}_2 \vee x_1 x_3) \vee x_1 x_2 \vee \overline{x}_1 \overline{x}_2 = \overline{x}_1 \overline{x}_2 \vee x_1 x_2 x_3 \vee x_1 x_2 \vee \overline{x}_1 \overline{x}_2.$$

Удаляя поглощаемые члены, завершаем преобразование

$$\overline{x}_1 \overline{x}_2 \vee x_1 x_2 x_3 \vee x_1 x_2 \vee \overline{x}_1 \overline{x}_2 = \overline{x}_1 \overline{x}_2 \vee x_1 x_2.$$

В результате тождественных преобразований была получена формула, в которую не входит переменная x_3. Поэтому x_3 является фиктивной переменной.

Данную выкладку можно было бы продолжить, написав в качестве окончательной формулы $x_1 \sim x_2$. Можно, однако, остановиться и на формуле $\overline{x}_1 \overline{x}_2 \vee x_1 x_2$, получая запись булевой функции в общепринятой стандартной форме. Подобные формулы, состоящие из конъюнкций переменных и их отрицаний (логических произведений), соединенных функциональными символами дизъюнкции, называются *дизъюнктивными нормальными формами* (д. н. ф.). Они являются основным средством представления булевых функций в математической логике. Далее будет показано, что любая, не равная тождественно нулю булева функция может быть выражена с помощью д. н. ф.

Для булевых функций существует и другая стандартная форма представления. Применив к полученной д. н. ф. закон дистрибутивности дизъюнкции по отношению к конъюнкции, имеем

$$\overline{x}_1 \overline{x}_2 \vee x_1 x_2 = (\overline{x}_1 \overline{x}_2 \vee x_1)(\overline{x}_1 \overline{x}_2 \vee x_2) =$$

$$= (\overline{x}_1 \vee x_1)(\overline{x}_2 \vee x_1)(\overline{x}_1 \vee x_2)(\overline{x}_2 \vee x_2) = (x_1 \vee \overline{x}_2)(\overline{x}_1 \vee x_2).$$

Заметим, к полученной формуле можно прийти быстрее, если воспользоваться свойством (4)

$$\overline{x}_1 \overline{x}_2 \vee x_1 x_2 = (\overline{x}_1 \overline{x}_2 \vee x_1)(\overline{x}_1 \overline{x}_2 \vee x_2) = (x_1 \vee \overline{x}_2)(\overline{x}_1 \vee x_2).$$

Полученное булево выражение, состоящее из дизъюнкций переменных и их отрицаний, соединенных функциональными символами конъюнкций, называется *конъюнктивной нормальной формой* (к. н. ф.). К. н. ф. является универсальным способом представления булевых функций, не равных тождественно единице.

Теперь, научившись преобразовывать произвольную булеву формулу в д. н. ф., можно предложить более систематический подход к решению рассмотренной в предыдущем разделе задачи, в которой требовалось сделать вывод о виновности Смита. Пусть P, Q, R и S обозначают те же высказывания, что и прежде. Возьмём конъюнкцию всех посылок и $\bar{Q}$

$$(P \to (Q \oplus R)) \,\&\, (\bar{Q} \to PS) \,\&\, (S \to (Q \oplus R)) \,\&\, \bar{Q}.$$

Если данная конъюнкция *выполнима*, т. е. существует набор значений переменных P, Q, R и S, обращающий её в единицу, то вывод о том, что Смит убийца, логически неправомерен. В противном же случае, при тождественном равенстве данной конъюнкции нулю, вывод о виновности Смита является логическим следствием посылок.

Для проверки выполнимости преобразуем каждую из посылок в дизъюнктивную нормальную форму, а затем, раскрыв скобки, преобразуем в дизъюнктивную нормальную форму всё булево выражение:

$$\begin{aligned}(P\to(Q\oplus R))(\bar{Q}\to PS)(S\to(Q\oplus R))\bar{Q} &= (\bar{P}\vee(Q\oplus R))(Q\vee PS)(\bar{S}\vee(Q\oplus R))\bar{Q} = \\ &= (\bar{P}\vee Q\bar{R}\vee\bar{Q}R)(Q\vee PS)(\bar{S}\vee Q\bar{R}\vee\bar{Q}R)\bar{Q} = \\ &= ((\bar{P}\vee Q\bar{R}\vee\bar{Q}R)(Q\vee PS))((\bar{S}\vee Q\bar{R}\vee\bar{Q}R)\bar{Q}) = \\ &= (\bar{P}Q\vee Q\bar{R}\vee PQ\bar{R}S\vee P\bar{Q}RS)(\bar{Q}\bar{S}\vee\bar{Q}R) = P\bar{Q}RS.\end{aligned}$$

Теперь сразу видно, что конъюнкция выполнима при $P=1$, $R=1$, $S=1$, $Q=0$. Таким образом, на основании имеющихся посылок нельзя сделать вывод, что Смит убийца, и (в соответствии с презумпцией невиновности) он должен быть признан невиновным.

Если же третья посылка имеет вид $S \to (\bar{Q} \oplus R)$, то конъюнкция посылок и $\bar{Q}$ становится тождественно равной нулю:

$$\begin{aligned}(P\to(Q\oplus R))(\bar{Q}\to PS)(S\to(\bar{Q}\oplus R))\bar{Q} &= (\bar{P}\vee(Q\oplus R))(Q\vee PS)(\bar{S}\vee(\bar{Q}\oplus R))\bar{Q} = \\ &= (\bar{P}\vee Q\bar{R}\vee\bar{Q}R)(Q\vee PS)(\bar{S}\vee\bar{Q}\bar{R}\vee QR)\bar{Q} = \\ &= ((\bar{P}\vee Q\bar{R}\vee\bar{Q}R)(Q\vee PS))((\bar{S}\vee\bar{Q}\bar{R}\vee QR)\bar{Q}) = \\ &= (\bar{P}Q\vee Q\bar{R}\vee PQ\bar{R}S\vee P\bar{Q}RS)(\bar{Q}\bar{S}\vee\bar{Q}\bar{R}) = 0.\end{aligned}$$

Теперь вывод о виновности Смита логически вытекает из посылок.

Хотя используемый метод проверки правильности логического следствия из заданного множества посылок и является вполне регулярным, потребовавшиеся для его осуществления выкладки оказались достаточно длинными. Это типично для решения задачи о выполнимости конъюнктивных форм, которая является трудоёмкой задачей. Подробнее это будет обсуждаться в Главе 4.

Вопросы для самопроверки

1. Какую логическую операцию нужно подставить вместо $*$, чтобы было справедливо равенство $x_1\bar{x}_2 \vee \bar{x}_1 x_2 = x_1 * x_2$?

   а) $\vee$; б) $\&$; в) $\oplus$.

Ответы: 1 — в.

2.3. Булевы функции и схемы

Булевы функции являются удобным средством для описания функционирования проводящих схем. На это обстоятельство в 1938 году впервые обратил внимание выдающийся американский математик и инженер Клод Шеннон (1916–2001). Пусть схема образована релейными устройствами замыкающего и размыкающего типа. Каждое реле управляется подаваемым на него сигналом. Замыкающее реле разомкнуто при отсутствии сигнала и замыкается при его наличии. Размыкающее реле замкнуто при отсутствии сигнала и размыкается, когда он подается. Замыкающему реле ставится в соответствие переменная x_i, размыкающему — $\bar{x}_i$, где $x_i = 1$ при наличии i-го сигнала и $x_i = 0$ при его отсутствии.

Пусть имеется представленная на рис. 1 релейная схема.

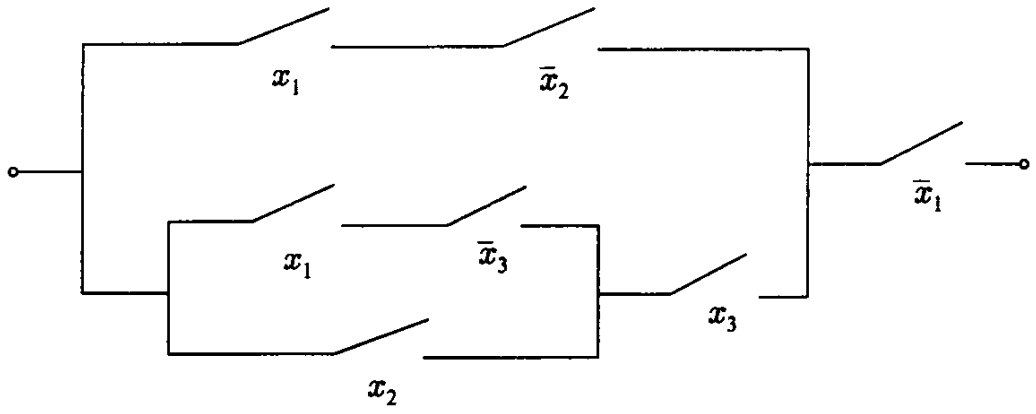


Рис. 1

Последовательное соединение реле описывается, как нетрудно понять, конъюнкцией соответствующих переменных, а параллельное — дизъюнкцией. Поэтому проводимость данной схемы описывается булевой формулой $((x_1\bar{x}_3 \vee x_2)x_3 \vee x_1\bar{x}_2)\bar{x}_1$, которая после упрощения приводится к виду $\bar{x}_1 x_2 x_3$. Таким образом, схема проводит в том и только в том случае, если отсутствует первый сигнал и присутствуют второй и третий.

В качестве более актуального, с точки зрения современных компьютерных технологий, приложения булевых функций рассмотрим схемы из функциональных элементов. Такую схему можно представлять себе как блок, на

вход которого поступают булевы переменные $x_1,\ldots,x_n$, а на выходе возникают функции от этих переменных $f_1,\ldots,f_m$ (рис. 2).

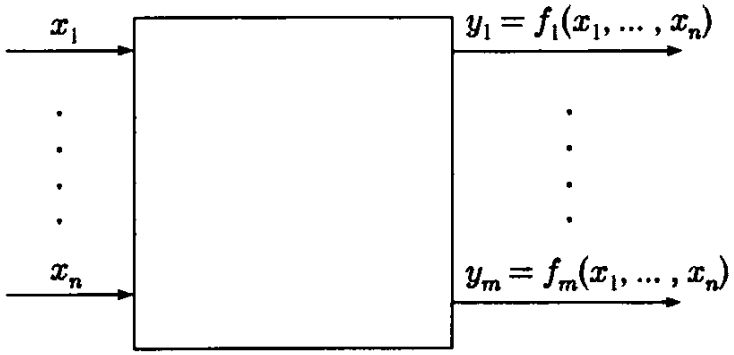
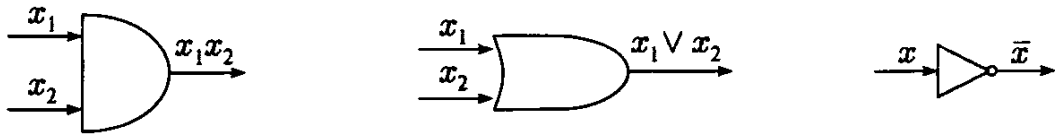


Рис. 2

Функции $f_1,\ldots,f_m$ реализуются в этом блоке с помощью стандартных функциональных элементов. Каждый такой элемент имеет несколько входов и один выход, на котором реализуется определённая булева функция. Этот выход может использоваться в качестве входа для других элементов или быть выходом схемы, реализующим одну из функций f_i. Считается, что выход любого элемента схемы при подаче набора значений переменных на её вход возникает мгновенно, без временной задержки. Такие схемы называют также *комбинационными схемами*. Наиболее часто в качестве стандартных функциональных элементов используются конъюнкторы, дизъюнкторы и инверторы. Их принятые в научно-технической литературе стандартные обозначения показаны на рис. 3.



Конъюнктор Дизъюнктор Инвертор

Рис. 3

Рассмотрим в качестве примера сложение двух одноразрядных двоичных чисел x_1 и x_2, которое, приводит, вообще говоря, к двухразрядному двоичному числу, так как единица может перенестись в старший разряд. Поэтому данная схема имеет два входа x_1 и x_2 и два выхода y_1 и y_2, где y_1 — младший, а y_2 — старший разряды полученной суммы. Таблицы истинности для функций y_1 и y_2 представлены в табл. 1.

x_1	x_2	y_1	y_2
0	0	0	0

0	1	1	0
1	0	1	0
1	1	0	1

Таблица 1

Как нетрудно заметить, $y_1 = x_1 \oplus x_2 = x_1\bar{x}_2 \vee \bar{x}_1 x_2 = (x_1 \vee x_2)\overline{x_1 x_2}$, $y_2 = x_1 x_2$. Поэтому данная операция может быть выполнена с помощью схемы на рис. 4, которая называется *полусумматором*, так как не учитывает возможной единицы, перенесённой из предыдущего разряда при сложении многоразрядных двоичных чисел.

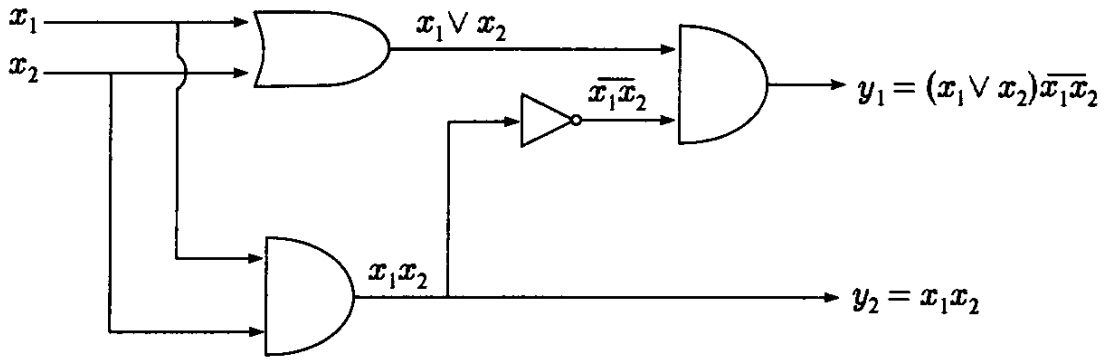


Рис. 4

Для краткости обозначим схему полусумматора как показано на рис. 5.

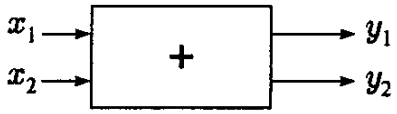


Рис. 5

Схема, называемая *полным сумматором*, выполняет сложение трёх одноразрядных двоичных чисел. Она реализует, таким образом, сложение одноимённых разрядов двух двоичных чисел с учётом того, что в результате сложения предыдущих разрядов в этот разряд могла быть перенесена единица. Входом для этой схемы являются три одноразрядных двоичных числа x_1, x_2, и x_3, где x_3 — результат переноса, а выходом, как и прежде, — два числа y_1 и y_2. Таблицы истинности для y_1 и y_2 представлены в табл. 2.

Так как $y_1 = (x_1 \oplus x_2) \oplus x_3$, то y_1 легко получить с помощью двух полусумматоров. Для y_2, как легко проверить, справедливо представление $y_2 = x_1 x_2 \vee x_1 x_3 \vee x_2 x_3 = x_1 x_2 \vee (x_1 \vee x_2) x_3$. Поэтому сумматор может быть реализован следующим образом (рис. 6).

x_1	x_2	x_3	y_1	y_2
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

Таблица 2

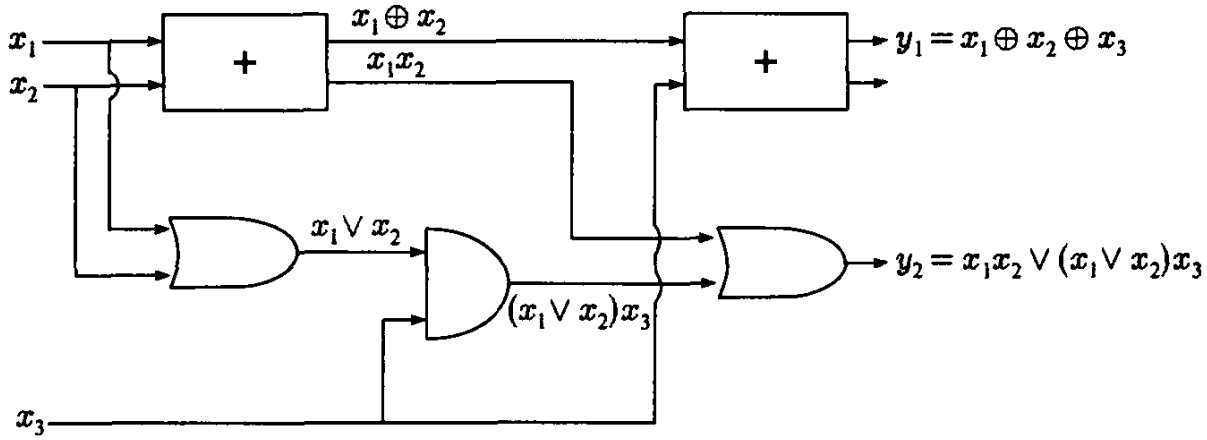
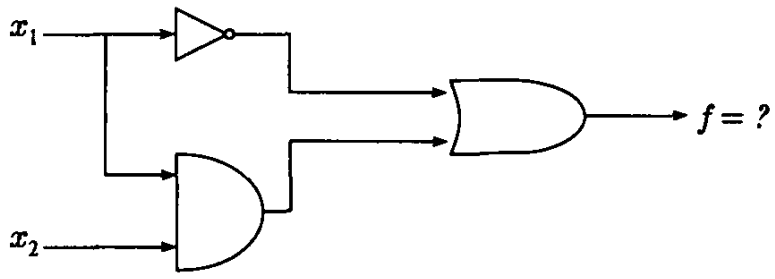


Рис. 6

Имея схему полного сумматора, легко построить схему и для сложения n -разрядных двоичных чисел.

Вопросы для самопроверки

1. Какую булеву функцию реализует данная схема?



а) $\overline{x}_1 x_2$; б) $x_1 x_2 \vee \overline{x}_1$; в) $\overline{x}_1 \vee x_2$.

Ответы: 1 — б.

2.4. Дизъюнктивная и конъюнктивная нормальные формы

Для дальнейшего нам будет удобно следующее обозначение. Пусть $\sigma \in \{0,1\}$, x — булева переменная. Тогда, по определению, $x^{\sigma} = x$, если $\sigma = 1$, и $x^{\sigma} = \overline{x}$, если $\sigma = 0$. Таким образом,

$$x^{\sigma} = 1 \Leftrightarrow x = \sigma,\ x^{\sigma} = 0 \Leftrightarrow x = \overline{\sigma}.$$

В дальнейшем также в качестве общего названия переменной с отрицанием или без отрицания будет использоваться термин *литерал*.

Выражение $x_{i_1}^{\sigma_1} x_{i_2}^{\sigma_2} \ldots x_{i_r}^{\sigma_r}$ называется элементарной конъюнкцией ранга r.

Так как в элементарной конъюнкции для каждой из n переменных есть три возможности: входить с отрицанием, входить без отрицания и не входить, то всего имеется 3^n элементарных конъюнкций. Пустая конъюнкция, не содержащая ни одного литерала, считается конъюнкцией нулевого ранга и полагается равной константе 1.

Выражение $x_{i_1}^{\sigma_1} \vee x_{i_2}^{\sigma_2} \vee \ldots \vee x_{i_r}^{\sigma_r}$ называется элементарной дизъюнкцией ранга r.

Пустая дизъюнкция считается дизъюнкцией нулевого ранга и полагается равной константе 0.

Следующие утверждения являются очевидным следствием введённых соглашений.

Утверждение 1. Элементарная конъюнкция $x_1^{\sigma_1} x_2^{\sigma_2} \ldots x_n^{\sigma_n}$ принимает значение 1 на единственном наборе значений переменных $x_1 = \sigma_1, x_2 = \sigma_2, \ldots, x_n = \sigma_n$.

Утверждение 2. Элементарная дизъюнкция $x_1^{\sigma_1} \vee x_2^{\sigma_2} \vee \ldots \vee x_n^{\sigma_n}$ принимает значение 0 на единственном наборе значений переменных $x_1 = \overline{\sigma}_1, x_2 = \overline{\sigma}_2, \ldots, x_n = \overline{\sigma}_n$.

Из этих утверждений вытекают следующие теоремы.

Теорема 1. Для любой булевой функции $f(x_1,\ldots,x_n)$, не равной тождественно нулю, справедливо следующее представление

$$f(x_1,\ldots,x_n)= \bigvee_{\substack{(\sigma_1,\ldots,\sigma_n)\in\{0,1\}^n \\ f(\sigma_1,\ldots,\sigma_n)=1}} x_1^{\sigma_1}x_2^{\sigma_2}\ldots x_n^{\sigma_n},$$

которое называется совершенной дизъюнктивной нормальной формой (совершенной д. н. ф.) булевой функции $f(x_1,\ldots,x_n)$.

(От ограничения $f \neq 0$ можно избавиться, если считать, что пустая д. н. ф. задаёт функцию, тождественно равную нулю.)

Теорема 2. Для любой булевой функции $f(x_1,\ldots,x_n)$, не равной тождественно единице, справедливо следующее представление

$$f(x_1,\ldots,x_n)= \&_{\substack{(\sigma_1,\ldots,\sigma_n)\in\{0,1\}^n \\ f(\sigma_1,\ldots,\sigma_n)=0}} x_1^{\bar{\sigma}_1}\vee x_2^{\bar{\sigma}_2}\vee\ldots\vee x_n^{\bar{\sigma}_n},,$$

которое называется совершенной конъюнктивной нормальной формой (совершенной к. н. ф.) булевой функции $f(x_1,\ldots,x_n)$.

(От ограничения $f \neq 1$ можно избавиться, если считать, что пустая к. н. ф. задаёт функцию, тождественно равную единице.)

Рассмотрим эти представления на примере. Пусть булева функция $f(x_1,x_2,x_n)$ задана следующей таблицей истинности (табл. 1).

x_1	x_2	x_3	$f(x_1,x_2,x_3)$
0	0	0	1
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	0
1	1	1	1

Таблица 1

Заметим, что, выбрав раз и навсегда лексикографический порядок двоичных наборов, любую функцию можно компактно задать битовой строкой её значений. Булева функция от n переменных будет представ-

лена, таким образом, 2^n-мерным двоичным вектором. В частности, заданная табл. 1 функция $f(x_1, x_2, x_n)$ запишется как (10011101). Её совершенная д. н. ф.

$$f(x_1, x_2, x_3) = \bar{x}_1\bar{x}_2\bar{x}_3 \vee \bar{x}_1 x_2 x_3 \vee x_1\bar{x}_2\bar{x}_3 \vee x_1\bar{x}_2 x_3 \vee x_1 x_2 x_3,$$

а совершенная к. н. ф.

$$f(x_1, x_2, x_3) = (x_1 \vee x_2 \vee \bar{x}_3)(x_1 \vee \bar{x}_2 \vee x_3)(\bar{x}_1 \vee \bar{x}_2 \vee x_3).$$

(Напомним, что при построении совершенной д. н. ф. принимаются во внимание лишь единичные наборы функции, а при построении совершенной к. н. ф. — её нулевые наборы)

Теоремы 1 и 2 могут быть обобщены и записаны в форме разложения по k произвольным переменным. Приведем такое обобщение для теоремы 1, выбрав k первых переменных.

Теорема 1bis. Для любой булевой функции $f(x_1,\ldots,x_n)$ и любого k $(1 \le k \le n)$ справедливо представление

$$f(x_1,\ldots,x_n) = \bigvee_{(\sigma_1,\ldots,\sigma_k)\in\{0,1\}^k} x_1^{\sigma_1} x_2^{\sigma_2} \ldots x_k^{\sigma_k} f(\sigma_1,\ldots,\sigma_k, x_{k+1},\ldots,x_n).$$

Доказательство. Для каждого набора $(\alpha_1,\ldots,\alpha_n) \in \{0,1\}^n$ значений переменных $(x_1,\ldots,x_n)$ лишь одна из конъюнкций вида $x_1^{\sigma_1} x_2^{\sigma_2} \ldots x_k^{\sigma_k}$ отлична от нуля, а именно, конъюнкция $x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_k^{\alpha_k}$. Поэтому правая часть формулы принимает на наборе $(\alpha_1,\ldots,\alpha_n)$ значение $f(\alpha_1,\ldots,\alpha_n)$, что и доказывает теорему. □

Наиболее часто данное разложение используется при $k = 1$, когда оно принимает вид

$$f(x_1,\ldots,x_n) = x_1 f(1, x_2,\ldots,x_n) \vee \bar{x}_1 f(0, x_2,\ldots,x_n)$$

и называется разложением по переменной x_1. Такое разложение можно получить из совершенной д. н. ф., вынося за скобки x_1 и $\bar{x}_1$. В нашем случае

$$f(x_1, x_2, x_3) = x_1(\bar{x}_2\bar{x}_3 \vee \bar{x}_2 x_3 \vee x_2 x_3) \vee \bar{x}_1(\bar{x}_2\bar{x}_3 \vee x_2 x_3).$$

Минимизация д. н. ф. В совершенных д. н. ф. и к. н. ф. все элементарные конъюнкции и дизъюнкции имеют одинаковый ранг, равный числу переменных n. Но если отказаться от этого требования, то часто удается упростить представление. Такие упрощения могут оказаться полезными как при технической реализации булевой функции, так и при теоретическом анализе её свойств.

Рассмотрим этот вопрос для д. н. ф. Число элементарных конъюнкций, входящих в д. н. ф., принято называть *длиной* д. н. ф.. Существуют два критерия сложности д. н. ф.: её длина и суммарное число присутствующих в ней литералов.

Дизъюнктивная нормальная форма, реализующая заданную булеву функцию, называется её кратчайшей д. н. ф., если она имеет минимальную длину среди всех д. н. ф., реализующих данную функцию, и минимальной — если минимально общее число входящих в д. н. ф. литералов (минимальна сумма рангов входящих в д. н. ф. элементарных конъюнкций).

Элементарная конъюнкция K называется допустимой конъюнкцией для функции $f(x_1,\ldots,x_n)$ или её импликантом, если на каждом наборе значений переменных, на котором конъюнкция K обращается в 1, функция f также принимает единичное значение, т. е. из $K=1$ следует $f=1$ ($K=1 \Rightarrow f=1$). Импликант K функции f называется её простым импликантом, если после удаления из K любого присутствующего в нём литерала возникает элементарная конъюнкция, не являющаяся импликантом функции f.

Множество тех наборов, на которых элементарная конъюнкция $K=x_{i_1}^{\sigma_1}x_{i_2}^{\sigma_2}\ldots x_{i_r}^{\sigma_r}$ принимает единичное значение, будем называть *интервалом,* соответствующим конъюнкции K, и обозначать N_K. Входящие в интервал N_K двоичные наборы получаются приписыванием переменным $x_{i_1},\ldots,x_{i_r}$ значений $x_{i_1}=\sigma_1,\ldots,x_{i_r}=\sigma_r$ и приписываниями произвольных булевых значений остальным $n-r$ переменным, поэтому интервал N_K, соответствующий конъюнкции ранга r, в случае n переменных имеет мощность $|N_K|=2^{n-r}$. Будем говорить, что это интервал *размерности* $n-r$.

Интервалы, соответствующие импликантам функции f, будем называть *допустимыми для f интервалами.* Они являются подмножествами множества N_f. При этом простым импликантам, соответствуют допустимые интервалы максимальной размерности. Задача построения д. н. ф. для данной булевой функции сводится, таким образом, к покрытию множества N_f допустимыми интервалами. При этом при построении простейших д. н. ф. используются интервалы, соответствующие простым импликантам.

Строго говоря, лишь при построении минимальной д. н. ф. условие простоты используемых импликантов является обязательным. При построении кратчайшей д. н. ф. могут быть, в принципе, использованы и импликанты, не являющиеся простыми. Но в кратчайшей д. н. ф. их всегда можно заменить простыми, получающимися из использованных в д. н. ф. удалением некоторых из литералов. Поэтому как минимальные, так и кратчайшие д. н. ф. строят из простых импликантов. Ясно, что дизъюнкция всех простых импликантов функции f является д. н. ф., представляющей функцию f. Эта д. н. ф. имеет специальное название.

Дизъюнкция всех простых импликантов функции f называется сокращённой д. н. ф. функции f.

Задача построения простейшей д. н. ф. решается, таким образом, в два этапа. На первом этапе строятся все простые импликанты, а на втором из них отбираются импликанты для простейшей д. н. ф. Кратчайшая и минимальная д. н. ф. булевой функции получаются, таким образом, из её сокращённой д. н. ф. удалением некоторых её простых импликантов. Рассмотрим сначала первый этап.

Построение сокращённой д. н. ф. булевой функции по её совершенной д. н. ф. может быть выполнено методом последовательного склеивания конъюнкций по переменным. Если в д. н. ф. имеются две конъюнкции одинакового ранга r вида $x_i K$ и $\bar{x}_i K$, то результатом их склеивания по переменной x_i является конъюнкция K ранга $r-1$.

На первом шаге алгоритма осуществляются все возможные попарные склеивания конъюнкций ранга n сокращённой д. н. ф. Одна и та же конъюнкция может участвовать в нескольких склейках ввиду выполняющегося в булевой алгебре тождества $K \vee K = K$. Если какая-либо конъюнкция не может быть склеена ни с какой другой, то она отбирается в качестве простого импликанта ранга n. После этого все оставшиеся конъюнкции ранга n, т. е. участвовавшие в склейках, удаляются.

На втором шаге рассматриваются только полученные в результате склеивания на первом шаге конъюнкции ранга $n-1$, и к ним применяется операция склеивания. Снова, если какая-либо конъюнкция не участвует ни в одной из склеек, то она отбирается в качестве простого импликанта ранга $n-1$. Затем оставляются только полученные в результате склеивания конъюнкции ранга $n-2$.

Процесс продолжается до тех пор, пока на некотором шаге i ни к одной из имеющихся конъюнкций ранга $n-i+1$ не может быть применена операция склеивания. Все эти конъюнкции отбираются в качестве простых импликантов ранга $n-i+1$, и процесс построения сокращённой д. н. ф. завершается.

Обратившись к совершенной д. н. ф. из приведённого примера, видим, что к первой и третьей конъюнкциям может быть применена операция

склеивания по переменной x_1, ко второй и пятой — также по переменной x_1, к третьей и четвертой — по переменной x_3 и, наконец, к четвертой и пятой — по переменной x_2. Таким образом, получаем 4 конъюнкции ранга 2: $\bar{x}_2\bar{x}_3$, x_2x_3, x_1x_3, $x_1\bar{x}_2$, к которым операция склеивания уже не применима. Следовательно, все они являются простыми импликантами, и сокращённая д. н. ф. принимает вид

$$f(x_1,x_2,x_3)=\bar{x}_2\bar{x}_3 \vee x_2x_3 \vee x_1x_3 \vee x_1\bar{x}_2 .$$

Задачу получения из сокращённой д. н. ф. функции её минимальной и кратчайшей д. н. ф. удобно представить с помощью таблицы, строки которой соответствуют единичным наборам функции f, а столбцы — её простым импликантам. В клетках таблицы ставится значение 1, если простой импликант принимает на данном наборе значение 1, и 0 — в противном случае. Такую таблицу называют таблицей Куайна по имени американского логика и философа Уилларда Куайна (Willard Van Orman Quine (1908–2000)). Среди логических проблем, которыми он занимался, была и задача упрощения д. н. ф.

В каждой строке таблицы Куайна содержится хотя бы одна единица. Это выражает тот факт, что д. н. ф., образованная всеми простыми импликантами, представляет функцию f. Считая, что каждый столбец покрывает те строки, в которых он содержит единицу, данную задачу можно сформулировать как задачу о покрытии строк столбцами. Тогда при построении кратчайшей д. н. ф. ищется покрытие строк минимальным числом столбцов, а при построении кратчайшей д. н. ф. — покрытие, минимальное по суммарному весу использованных столбцов, где в качестве веса столбца фигурирует ранг соответствующего ему простого импликанта. В нашем примере таблица Куайна принимает вид (табл. 2):

$x_1x_2x_3$	$\bar{x}_2\bar{x}_3$	x_2x_3	x_1x_3	$x_1\bar{x}_2$
0 0 0	1	0	0	0
0 1 1	0	1	0	0
1 0 0	1	0	0	1
1 0 1	0	0	1	1
1 1 1	0	1	1	0

Таблица 2

Как видно из таблицы, первая строка покрывается только первым столбцом, а вторая — только вторым. Поэтому первый и второй столбец обязательно должны войти в покрытие. После того, как они включены в покрытие, непокрытой остаётся лишь четвёртая строка, которая может быть

покрыта третьим или четвёртым столбцами. Это и даёт две простейшие д. н. ф., представляющие функцию f:

$$D_1 = \bar{x}_2\bar{x}_3 \vee x_2x_3 \vee x_1x_3 \text{ и } D_2 = \bar{x}_2\bar{x}_3 \vee x_2x_3 \vee x_1\bar{x}_2 .$$

В рассматриваемом примере все простые импликанты имеют одинаковый ранг 2. Поэтому задачи построения из них минимальной и кратчайшей д. н. ф. совпадают. Каждая из двух построенных д. н. ф. является и минимальной, и кратчайшей. В общем случае, однако, множества минимальных и кратчайших д. н. ф. могут не только не совпадать, но даже и не пересекаться.

Выполненное построение можно наглядно проиллюстрировать на единичном кубе. Координаты его вершин являются двоичными наборами длины 3. Это позволяет считать булеву функцию заданной в его вершинах. Множество вершин куба обозначим $B^3 = \{0,1\}^3$. Вершины, на которых функция равна единице, будем называть *единичными,* а вершины, на которых функция равна нулю, — *нулевыми.* Интервалы, соответствующие конъюнкциям ранга 3 — это вершины куба, конъюнкциям ранга 2 — рёбра куба, конъюнкциям ранга 1 — грани куба.

На рис. 1 жирно выделены единичные вершины рассматриваемой функции и рёбра, соответствующие простым импликантам.

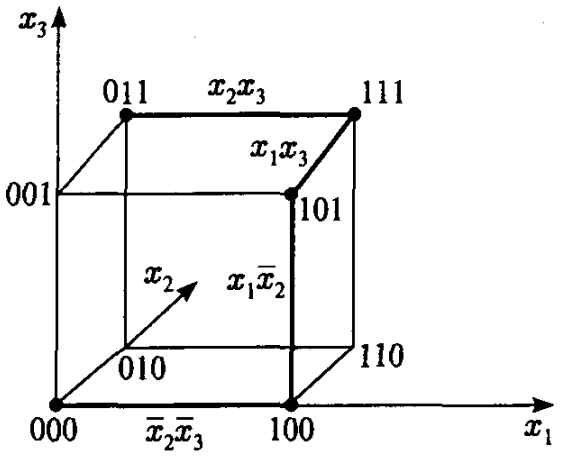


Рис. 1

Как видно из рис. 1, две полученные простейшие д. н. ф. соответствуют двум неизбыточным покрытиям единичных вершин интервалами, соответствующими простым импликантам, т. е. таким покрытиям, из которых удаление любого интервала приводит к появлению непокрытых единичных вершин.

Д. н. ф., реализующая функцию f и составленная из её простых импликантов, называется тупиковой д. н. ф. функции f, если при удалении из д. н. ф. любой входящей в неё конъюнкции полученная д. н. ф. реализует функцию, отличную от f.

Все минимальные д. н. ф. функции f и составленные из простых импликантов её кратчайшие д. н. ф. обязаны быть тупиковыми, но не каждая тупиковая д. н. ф. является минимальной или кратчайшей. Построение всех тупиковых д. н. ф. булевой функции по её сокращённой д. н. ф., т. е. синтез всех неизбыточных покрытий таблицы Куайна может быть осуществлен методом преобразования некоторой к. н. ф. в д. н. ф.

Для получения исходной к. н. ф. свяжем с j-м столбцом таблицы Куайна булеву переменную z_j, считая, что $z_j = 1$, если j-й столбец включается в покрытие, и $z_j = 0$, если не включается. Условие покрытия i-й строки, выражается дизъюнкцией тех переменных z_j, которые соответствуют покрывающим i-ю строку столбцам. А условие покрытия всех строк таблицы выражается конъюнкцией этих дизъюнкций по всем строкам.

Рассмотрим этот метод на примере табл. 2. Первая строка покрывается только первым столбцом, вторая — только вторым, третья — первым и четвёртым, четвёртая — третьим и четвёртым и пятая — вторым и третьим столбцами. Поэтому исходная к. н. ф. принимает вид

$$z_1 \& z_2 \& (z_1 \vee z_4) \& (z_3 \vee z_4) \& (z_2 \vee z_3).$$

Раскрывая скобки и удаляя поглощаемые члены, получаем выражение

$$z_1 z_2 z_3 \vee z_1 z_2 z_4,$$

в котором каждая элементарная конъюнкция соответствует тупиковой д. н. ф.

Вкратце рассмотрим ещё один элементарный пример функции от трёх переменных, заданной строкой своих значений как (10111010). Заинтересованный читатель самостоятельно восстановит её таблицу истинности и геометрически проиллюстрирует на единичном кубе. Её совершенная д. н. ф. запишется как

$$\bar{x}_1 \bar{x}_2 \bar{x}_3 \vee \bar{x}_1 x_2 \bar{x}_3 \vee \bar{x}_1 x_2 x_3 \vee x_1 \bar{x}_2 \bar{x}_3 \vee x_1 x_2 \bar{x}_3.$$

Склеивая 1-ю и 2-ю конъюнкции по x_2, 1-ю и 4-ю по x_1, 2-ю и 3-ю по x_3, 2-ю и 5-ю по x_1, 4-ю и 5-ю по x_2, получаем д. н. ф.

$$\bar{x}_1 \bar{x}_3 \vee \bar{x}_2 \bar{x}_3 \vee \bar{x}_1 x_2 \vee x_2 \bar{x}_3 \vee x_1 \bar{x}_3.$$

В ней выполнимы 2 склейки: 1-ю и 5-ю по x_1, 2-ю и 4-ю по x_2, которые приводят к одной и той же конъюнкции $\bar{x}_3$, являющейся простым импликантом. Другим простым импликантом является не участвовавшая в склейках 3-я конъюнкция. Таким образом, сокращённая д. н. ф. имеет вид

$$\bar{x}_3 \vee \bar{x}_1 x_2.$$

Так как ни один из её простых импликантов не может быть удалён, то она является тупиковой, кратчайшей и минимальной д. н. ф.

Рассмотренные элементарные примеры минимизации функций, зависящих от трёх переменных, не дают в полной мере представления о тех трудностях, которые возникают при построении простейших д. н. ф. для функций, зависящих от большего числа переменных. Поэтому, завершая данный раздел, сделаем ряд заключительных замечаний, призванных правильно сориентировать читателя в этих вопросах.

1. При большом числе переменных число простых импликантов булевой функции, как правило, значительно превышает число её единичных вершин. Другими словами, длина сокращённой д. н. ф. оказывается больше длины совершенной д. н. ф. (см. Дополнение 2). Термин «сокращённая д. н. ф.» нельзя, таким образом, признать удачным. Его появление объясняется недостаточной изученностью вопросов минимизации д. н. ф. на раннем этапе развития теории в пятидесятые годы прошлого века.
2. Число тупиковых д. н. ф. функции $f(x_1,\ldots,x_n)$ для большинства функций быстро растёт с ростом n (см. задачу 9) и даже может достигать $(2^{2^n})^{cn}$, где c — некоторая константа [18]. Поэтому метод минимизации, основанный на построении всех тупиковых д. н. ф. и выборе из этого множества простейшей д. н. ф., нерационален. При большом числе переменных он может оказаться неосуществимым даже при наличии мощного компьютера.
3. Задача о покрытии строк таблицы Куайна её столбцами, которую приходится решать при построении кратчайшей и минимальной д. н. ф., является трудной задачей дискретной оптимизации, точное решение которой даже на компьютере возможно лишь для таблиц небольшой размерности. Поэтому при большой размерности задачи приходится ограничиваться приближённым решением. (Задача о покрытии в общей постановке обсуждается в Главе 4)
4. Склеивание элементарных конъюнкций не есть единственный способ получения сокращённой д. н. ф. Сокращённую д. н. ф. для таблично заданной булевой функции можно получить, выписав её совершенную к. н. ф. и преобразовав её в д. н. ф. Если по ходу преобразования удалять поглощаемые конъюнкции, то полученная д. н. ф. будет сокращённой. Этот метод, часто называемый алгоритмом Нельсона, не является более экономным, но он поучителен. В нашем примере имеем

$$f(x_1,x_2,x_3)=(x_1\vee x_2\vee\bar{x}_3)(x_1\vee\bar{x}_2\vee x_3)(\bar{x}_1\vee\bar{x}_2\vee x_3)=$$

$$=(x_1x_1\vee x_1\bar{x}_2\vee x_1x_3\vee x_1x_2\vee x_2\bar{x}_2\vee x_2x_3\vee x_1\bar{x}_3\vee\bar{x}_2\bar{x}_3\vee x_3\bar{x}_3)(\bar{x}_1\vee\bar{x}_2\vee x_3)=$$

$$=(x_1\vee x_2x_3\vee\bar{x}_2\bar{x}_3)(\bar{x}_1\vee\bar{x}_2\vee x_3)=x_1\bar{x}_2\vee x_1x_3\vee\bar{x}_1x_2x_3\vee x_2x_3\vee\bar{x}_1\bar{x}_2\bar{x}_3\vee\bar{x}_2\bar{x}_3=$$

$$=x_1\bar{x}_2\vee x_1x_3\vee x_2x_3\vee\bar{x}_2\bar{x}_3.$$

5. При нахождении сокращённой д. н. ф. булева функция не обязательно должна быть изначально представлена своей совершенной д. н. ф. или совершенной к. н. ф. Метод Блейка позволяет строить сокращённую д. н. ф. для булевой функции, заданной любой своей д. н. ф. (см. задачу 10).
6. При $n > 3$ булеву функцию уже нельзя столь наглядно представить с помощью единичного куба B^3. Однако геометрическую терминологию часто сохраняют и в общем случае, считая булеву функцию заданной на вершинах единичного n-мерного куба (гиперкуба), который называют также *булевым кубом* и обозначают $B^n = \{0,1\}^n$. Его вершинами являются двоичные наборы длины n. Рёбра куба соединяют вершины, отличающиеся в одной координате. В дальнейшем термины «двоичный набор длины n» и «вершина единичного n-мерного куба» будут синонимами. Для них в соответствии с отечественной традицией могут использоваться обозначения $\tilde{x} = (x_1, \ldots, x_n)$ и $\bar{\tilde{x}} = (\bar{x}_1, \ldots, \bar{x}_n)$. На рис. 2 схематически представлен четырёхмерный куб B^4, получающийся из трёхмерного «сдвигом» по четвёртой координате.

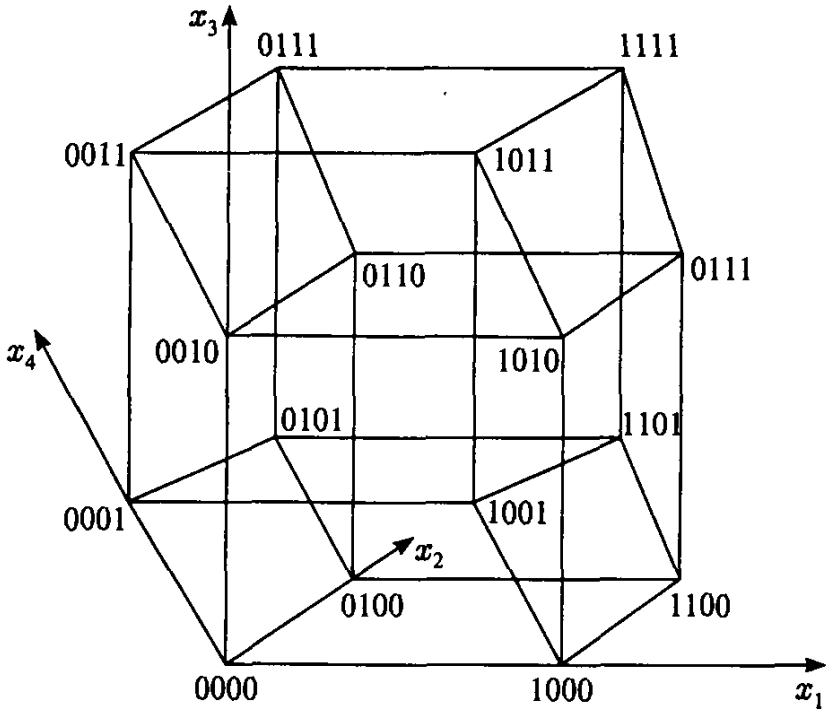


Рис. 2

7. В практических научно-технических задачах часто приходится минимизировать не всюду определённую, или, как говорят, *частичную булеву функцию*. В этом случае заданные единичные и нулевые наборы не исчерпывают собой множества двоичных наборов переменных, и на некоторых из наборов функция остаётся неопределённой.

(Этот случай возможен, например, когда некоторые из наборов значений переменных не возникают при функционировании реализующего булеву функцию технического устройства) При реализации частичной булевой функции с помощью д. н. ф. осуществляется её доопределение, позволяющее получить простейшую д. н. ф. Принципиально алгоритм минимизации при этом не изменяется. На первом шаге находятся максимальные допустимые интервалы, т. е. такие конъюнкции K минимально возможного ранга, для которых справедливо $K=1 \Rightarrow f \neq 0$. Затем на втором шаге из этих конъюнкций строится минимальное покрытие для множества единичных наборов частичной булевой функции.

Докажем теперь общее утверждение, относящееся к максимально возможной длине кратчайшей д. н. ф. булевой функции.

Теорема 3. Максимум длины кратчайшей д. н. ф. булевой функции от n переменных равен 2^{n-1}. Он достигается, например, на функции $f(x_1,\ldots,x_n)=x_1 \oplus x_2 \oplus \ldots \oplus x_n$, называемой счётчиком чётности.

Доказательство. Покажем, что произвольную булеву функцию $f(x_1,\ldots x_n)$ можно записать с помощью д. н. ф., длина которой не превосходит 2^{n-1}. Рассмотрим 2^{n-1} конъюнкций вида $x_1^{\sigma_1}\ldots x_{n-1}^{\sigma_{n-1}}$, где $(\sigma_1,\ldots,\sigma_{n-1}) \in \{0,1\}^{n-1}$. Соответствующие им одномерные интервалы являются рёбрами куба. Они, как легко видеть, покрывают множество вершин куба. Строим теперь реализующую функцию д. н. ф. следующим образом. Если обе вершины ребра единичные, то включаем соответствующую ему конъюнкцию ранга $n-1$ в д. н. ф. Если единичной является лишь одна из вершин ребра, то включаем в д. н. ф. соответствующую ей конъюнкцию ранга n. Если обе вершины ребра нулевые, то пропускаем это ребро, ничего в д. н. ф. не включая. Полученная таким образом д. н. ф. реализует $f(x_1,\ldots x_n)$, и её длина не превышает 2^{n-1}.

Интервал, соответствующий произвольной конъюнкции, является связным множеством, т. е. из любой его вершины можно попасть в любую другую, переходя каждый раз по ребру в соседнюю вершину и не покидая пределов интервала. Счётчик чётности имеет 2^{n-1} изолированных единичных вершин. Поэтому никакие две из них нельзя покрыть допустимым для этой функции интервалом, и для покрытия каждой единичной вершины требуется отдельная конъюнкция. Таким образом, никакие две конъюнкции совершенной д. н. ф. не могут быть склеены, и длина кратчайшей д. н. ф. счётчика чётности совпадает с длиной его совершенной д. н. ф., т. е. равна 2^{n-1}. □

Вопросы для самопроверки

1. Какая из трёх д. н. ф. реализует функцию $x_1 \oplus x_2 \oplus 1$?

   а) $x_1\bar{x}_2 \vee \bar{x}_1 x_2$; б) $x_1 \vee x_2 \vee x_1 x_2$; в) $x_1 x_2 \vee \bar{x}_1\bar{x}_2$.

2. Какая из трёх к. н. ф. реализует функцию $x_1 \oplus x_2 \oplus 1$?

   а) $(\bar{x}_1 \vee x_2)(x_1 \vee \bar{x}_2)$; б) $(x_1 \vee x_2)(\bar{x}_1 \vee \bar{x}_2)$; в) $x_1 x_2 (x_1 \vee x_2)$.

Ответы: 1 — в, 2 — а.

2.5. Двойственность

Если взглянуть на все те встречавшиеся до сих пор тождества булевой алгебры, куда входят только операции $\{\vee, \&, \neg\}$, то можно заметить, что операции $\vee$ и &, а также константы 0 и 1 входят в них равноправно и симметрично. Точнее, если в каком-либо из соотношений заменить все операции $\vee$ на &, а все операции & на $\vee$, одновременно заменив 1 на 0 и 0 на 1, то снова получится одно из выписанных соотношений (другое или то же самое). Например, друг в друга переходят правила де Моргана, а также законы дистрибутивности.

Подобная симметрия операций $\vee$ и & не является случайностью и находит объяснение в принципе двойственности. Он основан на перестановке местами констант 0 и 1, осуществляемой операцией отрицания. Если в булевой алгебре константы 1 и 0 поменять местами, т. е. заменить 0 на 1 и 1 на 0, то функция $f(x_1,\ldots,x_n)$ перейдет в функцию $\bar{f}(\bar{x}_1,\ldots,\bar{x}_n)$ (изменяются на противоположные как значения переменных, так и значение функции). Полученная в результате такой замены функция имеет специальное название и обозначение.

Функция $f^*(x_1,\ldots,x_n) = \bar{f}(\bar{x}_1,\ldots,\bar{x}_n)$ называется двойственной к функции $f(x_1,\ldots,x_n)$.

Утверждение 1. Соотношение двойственности является взаимным: $f^{**}(x_1,\ldots,x_n) = f(x_1,\ldots,x_n)$.

Доказательство.

$$f^{**}(x_1,\ldots,x_n) = \left(\bar{f}(\bar{x}_1,\ldots,\bar{x}_n)\right)^* = \bar{\bar{f}}(\bar{\bar{x}}_1,\ldots,\bar{\bar{x}}_n) = f(x_1,\ldots,x_n). \qquad \square$$

Утверждение 2. Дизъюнкция и конъюнкция двойственны друг другу: $(x_1 \vee x_2)^* = x_1 \& x_2$.

Доказательство. Доказательство следует из правил де Моргана:

$$(x_1 \vee x_2)^* = \overline{(\bar{x}_1 \vee \bar{x}_2)} = \bar{\bar{x}}_1 \,\&\, \bar{\bar{x}}_2 = x_1 \,\&\, x_2 .$$ □

Заметим, что штрих Шеффера и стрелка Пирса также являются двойственной парой:

$$(x_1 \mid x_2)^* = (x_1 \downarrow x_2) .$$

Утверждение 3. Функция, двойственная к суперпозиции булевых функций, есть суперпозиция двойственных функций:

$$(f(g_1(x_1,\ldots,x_n),\ldots,g_m(x_1,\ldots,x_n))^* =$$

$$= f^*(g_1^*(x_1,\ldots,x_n),\ldots,g_m^*(x_1,\ldots,x_n)) .$$

Доказательство. Доказательство следует из цепочки равенств:

$$\left(f(g_1(x_1,\ldots,x_n),\ldots,g_m(x_1,\ldots,x_n))\right)^* = \bar{f}(g_1(\bar{x}_1,\ldots,\bar{x}_n),\ldots,g_m(\bar{x}_1,\ldots,\bar{x}_n)) =$$

$$= \bar{f}(\bar{\bar{g}}_1(\bar{x}_1,\ldots,\bar{x}_n),\ldots,\bar{\bar{g}}_m(\bar{x}_1,\ldots,\bar{x}_n)) = \bar{f}\left(\overline{g_1^*}(x_1,\ldots,x_n),\ldots,\overline{g_m^*}(x_1,\ldots,x_n)\right) =$$

$$= f^*(g_1^*(x_1,\ldots,x_n),\ldots,g_m^*(x_1,\ldots,x_n)).$$ □

Теорема 1 (принцип двойственности). Если в некотором булевом тождестве, содержащем лишь операции $\{\vee, \&, \neg\}$, заменить все конъюнкции на дизъюнкции, а дизъюнкции — на конъюнкции, а также заменить 0 на 1 и 1 на 0, то снова возникнет некоторое тождество булевой алгебры.

Доказательство. Так как операции $\vee$ и $\&$ взаимно двойственны, то их замена друг на друга приведет в силу леммы 1 к тому, что обе части тождества заменятся соответственно на двойственные функции, которые также будут равны, и равенство сохранится. □

Подобными тождествами, получаемыми друг из друга с помощью принципа двойственности, являются правила де Моргана, а также законы дистрибутивности для дизъюнкции и конъюнкции.

В качестве ещё одного примера применения принципа двойственности рассмотрим равенство

$$x_1x_2 \vee \bar{x}_1\bar{x}_2 \vee x_2x_3 \vee \bar{x}_2\bar{x}_3 \vee x_1x_3 \vee \bar{x}_1\bar{x}_3 = 1 .$$

Как нетрудно проверить полным перебором всех наборов значений переменных, левая часть равна единице при любом наборе значений переменных, т. е. данное равенство является тождеством. (Подобные тожде-

ства в логике называются *тавтологиями*). С помощью принципа двойственности из него может быть получено другое тождество

$$(x_1 \vee x_2)(\overline{x}_1 \vee \overline{x}_2)(x_2 \vee x_3)(\overline{x}_2 \vee \overline{x}_3)(x_1 \vee x_3)(\overline{x}_1 \vee \overline{x}_3) = 0.$$

Продолжим изучение понятия двойственности.

Функция $f(x_1,\ldots,x_n)$ называется самодвойственной, если

$$f^*(x_1,\ldots,x_n) = f(x_1,\ldots,x_n).$$

Самодвойственность функции $f(x_1,\ldots,x_n)$ эквивалентна тому, что

$$f(\overline{x}_1,\ldots,\overline{x}_n) = \overline{f}(x_1,\ldots,x_n).$$

Утверждение 4. Имеется $2^{2^{n-1}}$ самодвойственных функций от n переменных.

Доказательство. На любом наборе из каждой пары двух взаимно противоположных наборов $(x_1,\ldots,x_n)$ и $(\overline{x}_1,\ldots,\overline{x}_n)$ самодвойственную функцию можно задать произвольно. Тогда на другом наборе пары функция однозначно определится, приняв противоположное значение. Так как имеется 2^{n-1} пар взаимно противоположных наборов, число самодвойственных функций равно $2^{2^{n-1}}$. □

Двумя самодвойственными функциями от одной переменной являются x и $\overline{x}$. Четырьмя самодвойственными функциями от двух переменных являются x_1, $\overline{x}_1$, x_2, $\overline{x}_2$, что свидетельствует об отсутствии самодвойственных функций, существенно зависящих от двух переменных. Имеется 10 самодвойственных функций, существенно зависящих от трёх переменных, например $x_1 \oplus x_2 \oplus x_3$ или $x_1x_2 \vee x_1x_3 \vee x_2x_3$.

В качестве непосредственного следствия из утверждения 3 получаем также следующий важный результат.

Утверждение 5. Функция, полученная путём суперпозиции из самодвойственных функций, является самодвойственной.

Вопросы для самопроверки

1. Какая из следующих функций двойственна функции $x_1\overline{x}_2\overline{x}_3$?

   а) $\overline{x}_1x_2x_3$; б) $x_1 \vee \overline{x}_2 \vee \overline{x}_3$; в) $\overline{x}_1 \vee x_2 \vee x_3$.

2. Сколько существует самодвойственных функций от трёх переменных?

   а) 2; б) 8; в) 16.

Ответы: 1 — б, 2 — в.

2.6. Геометрия единичного n-мерного куба

Метрика Хэмминга. Геометрическая интерпретация булевой функции $f(x_1,\ldots,x_n)$ на множестве вершин единичного n-мерного куба B^n подсказывает глубже рассмотреть его геометрию. Прежде всего, отметим, что интервал, соответствующий конъюнкции ранга r, изоморфен $(n-r)$-мерному кубу B^{n-r} и поэтому его часто называют *подкубом*. Рассмотрим теперь на множестве вершин гиперкуба B^n метрику (расстояние между вершинами), которая используется в самых различных задачах дискретного анализа и называется метрикой Хэмминга, в честь американского математика Ричарда Хэмминга (Richard Wesley Hamming (1915–1998)), специалиста в области вычислительной техники, одного из создателей теории помехоустойчивого кодирования.

Пусть $\tilde{\alpha},\tilde{\beta}\in B^n$ — две вершины единичного n-мерного куба. Расстоянием Хэмминга $d(\tilde{\alpha},\tilde{\beta})$ между ними называется число несовпадающих у них координат.

Например, если $\tilde{\alpha}=(0100110)$, $\tilde{\beta}=(1001100)$, то $d(\tilde{\alpha},\tilde{\beta})=4$, так как наборы отличаются по первой, второй, четвёртой и шестой координатам.

Расстояние Хэмминга равно длине кратчайшего пути между вершинами по рёбрам куба и удовлетворяет обычным аксиомам метрики:

1) *неотрицательности* — $d(\tilde{\alpha},\tilde{\beta})\geq 0$, *причём* $d(\tilde{\alpha},\tilde{\beta})=0$ *в том и только в том случае, если* $\tilde{\alpha}=\tilde{\beta}$;
2) *симметричности* — $d(\tilde{\alpha},\tilde{\beta})=d(\tilde{\beta},\tilde{\alpha})$;
3) *неравенству треугольника* — $d(\tilde{\alpha},\tilde{\beta})+d(\tilde{\beta},\tilde{\gamma})\geq d(\tilde{\alpha},\tilde{\gamma})$.

Вершины, находящиеся на расстоянии 1 друг от друга, называются *соседними*. Две соседние вершины образуют *ребро* гиперкуба. Последовательность соседних вершин $\tilde{\alpha}_1,\ldots,\tilde{\alpha}_k$ называется *цепью длины* $k-1$, связывающей вершины $\tilde{\alpha}_1$ и $\tilde{\alpha}_k$. Если $d(\tilde{\alpha}_1,\tilde{\alpha}_k)=k-1$, то эта цепь является кратчайшей цепью, связывающей вершины $\tilde{\alpha}_1$ и $\tilde{\alpha}_k$. Для любых двух вершин, находящихся на расстоянии l друг от друга существует $l!$ связывающих их кратчайших цепей. Множество принадлежащих им вершин изоморфно l-мерному кубу.

Дадим ещё ряд определений, которые в дальнейшем будут широко использоваться.

Весом $|\tilde{\alpha}|$ двоичного набора $\tilde{\alpha}$ называется число единиц в нём.

Вес вершины, таким образом, равен её расстоянию от нулевой вершины: $|\tilde{\alpha}| = d(\tilde{\alpha}, \tilde{0})$. Определив $\tilde{\alpha} \oplus \tilde{\beta}$ как набор, полученный из наборов $\tilde{\alpha}$ и $\tilde{\beta}$ их покомпонентным сложением по модулю 2, имеем также соотношение $d(\tilde{\alpha}, \tilde{\beta}) = |\tilde{\alpha} \oplus \tilde{\beta}|$.

Множество B_i^n вершин веса i называется i-м слоем единичного n-мерного куба $(i = 0,1,\ldots,n)$.

Таким образом, единичный n-мерный куб разбивается на $n+1$ слоёв $B^n = \bigcup_{i=0}^{n} B_i^n$, как показано на рис. 1.

Мощность i-го слоя $|B_i^n| = C_n^i$, а минимальное расстояние между вершинами одного слоя равно 2. Основная масса вершин сосредоточена в средних слоях куба. Точнее, справедливо следующее, вытекающее из (1.7.22) и (1.7.23) утверждение.

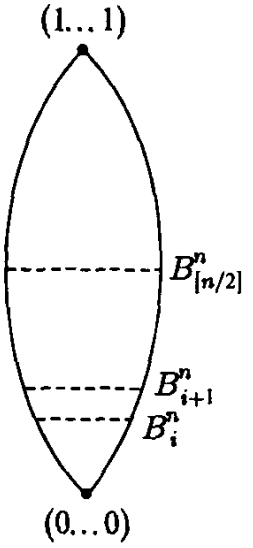


Рис. 1

Утверждение 1. Пусть $t(n)$ — произвольная сколь угодно медленно растущая функция, такая что $\lim_{n\to\infty} t(n) = \infty$. Тогда доля вершин, лежащих в слоях с номерами от $[n/2 - t\sqrt{n}]$ до $[n/2 + t\sqrt{n}]$, с ростом n стремится к единице.

Множество $S_r(\tilde{\alpha}) = \{\tilde{\beta} : d(\tilde{\alpha}, \tilde{\beta}) = r\}$ вершин, находящихся на расстоянии r от вершины $\tilde{\alpha}$, называется сферой Хэмминга радиуса r с центром в $\tilde{\alpha}$.

Мощность сферы Хэмминга радиуса r есть $|S_r(\tilde{\alpha})| = C_n^r$. Заметим, что $B_i^n = S_i(\tilde{0})$, т. е. i-й слой куба является сферой Хэмминга радиуса i с центром в начале координат.

Множество $\hat{S}_r(\tilde{\alpha}) = \bigcup_{i=0}^{r} S_r(\tilde{\alpha})$ вершин, находящихся на расстоянии не более чем r от вершины $\tilde{\alpha}$, называется шаром Хэмминга радиуса r с центром в $\tilde{\alpha}$.

Мощность шара Хэмминга радиуса r есть $|\widehat{S}_r(\tilde{\alpha})| = \sum_{i=0}^{r} C_n^i$. Эту часто встречающуюся в дискретной математике мощность будем в дальнейшем обозначать как $\chi(n,r) = |\widehat{S}_r(\tilde{\alpha})| = \sum_{i=0}^{r} C_n^i$.

Монотонные функции. Двоичные наборы — вершины куба, на которых задана булева функция, можно рассматривать и как характеристические векторы подмножеств n-элементного множества, считая тем самым булеву функцию заданной на этих подмножествах. Весом набора в этом случае является мощность соответствующего подмножества. При этом отношение включения подмножеств задаёт отношение частичного порядка на вершинах гиперкуба.

Для двух вершин куба $\tilde{\alpha}, \tilde{\beta} \in B^n$ отношение частичного порядка «$\geq$» определяется как

$$\tilde{\alpha} \leq \tilde{\beta} \Leftrightarrow \alpha_i \leq \beta_i, \quad i = 1,2,\ldots,n.$$

Как обычно, будем писать $\tilde{\alpha} < \tilde{\beta}$, если $\tilde{\alpha} \leq \tilde{\beta}$ и $\tilde{\alpha} \neq \tilde{\beta}$. Если $\tilde{\alpha} \leq \tilde{\beta}$ и $d(\tilde{\alpha}, \tilde{\beta}) = 1$, то будем говорить, что вершина $\tilde{\beta}$ *покрывает* вершину $\tilde{\alpha}$ и обозначать это как $\tilde{\beta} \succ \tilde{\alpha}$. Это означает, что набор $\tilde{\alpha}$ получается из набора $\tilde{\beta}$ заменой одной из его единичных координат на нулевую. Если для двух вершин (наборов) $\tilde{\alpha}$ и $\tilde{\beta}$ не выполнено ни одно из соотношений $\tilde{\alpha} \leq \tilde{\beta}$ или $\tilde{\beta} \leq \tilde{\alpha}$, то говорят, что они *несравнимы*.

Подмножество вершин куба называется антицепью, или шпернеровым семейством, если любые две вершины этого подмножества несравнимы.

Вершины слоя образуют, очевидно, антицепь. Максимальным по мощности среди слоев является средний слой n-мерного куба, содержащий $C_n^{[n/2]}$ вершин. При $n = 2k$ единственным максимальным слоем является k-й слой, при $n = 2k+1$ максимальных слоя два — k-й и $(k+1)$-й. В разделе 3.7 будет показано, что эти слои и являются единственными максимальными антицепями гиперкуба (лемма 3.7.1 (Шпернера)).

В ряде прикладных задач естественным образом возникают булевы функции, заданные на подмножествах некоторого множества и обладающие тем свойством, что если функция равна единице на некотором подмножестве, то она равна единице и на всех подмножествах, включающих данное.

Булева функция $f(\tilde{x})$ называется монотонной, если из $\tilde{\alpha} \le \tilde{\beta}$ следует $f(\tilde{\alpha}) \le f(\tilde{\beta})$.

Набор (вершина) $\tilde{\alpha} = (\alpha_1, \ldots, \alpha_n)$ называется нижней единицей монотонной функции $f(\tilde{x})$, если $f(\tilde{\alpha}) = 1$ и $f(\tilde{\beta}) = 0$ на всех наборах $\tilde{\beta}$ таких, что $\tilde{\beta} < \tilde{\alpha}$.

Ясно, что две различные нижние единицы монотонной булевой функции несравнимы между собой. Поэтому множество нижних единиц образует антицепь. Обратно, из любой антицепи можно получить монотонную булеву функцию, приняв вершины антицепи за нижние единицы функции. Таким образом, существует взаимно однозначное соответствие между антицепями и монотонными булевыми функциями.

Пусть $\tilde{\alpha} = (\alpha_1, \ldots, \alpha_n)$ — нижняя единица монотонной булевой функции $f(\tilde{x})$, $i_1, i_2, \ldots, i_r$ — номера единичных координат набора $\tilde{\alpha}$. Поставим в соответствие нижней единице $\tilde{\alpha}$ конъюнкцию $K_{\tilde{\alpha}} = x_{i_1} x_{i_2} \ldots x_{i_r}$. Конъюнкция $K_{\tilde{\alpha}}$ обращается в единицу на наборе $\tilde{\alpha}$ и всех больших наборах, поэтому $K_{\tilde{\alpha}}$ является импликантом функции $f(\tilde{x})$. Более того, она является простым импликантом, так как при отбрасывании любой буквы она, в силу определения нижней единицы, перестанет быть импликантом.

Рассмотрим теперь произвольный простой импликант монотонной булевой функции $f(\tilde{x})$. Он не может содержать переменных с отрицаниями, так как отбрасывание таких переменных привело бы, в силу монотонности функции, к допустимой конъюнкции, что противоречит определению простого импликанта. Ясно также, что соответствующий простому имликанту набор обязан быть нижней единицей функции. Поэтому все простые импликанты монотонной функции — суть конъюнкции, соответствующие её нижним единицам.

Заметим, наконец, что ни один из простых импликантов $K_{\tilde{\alpha}}$ не может быть удалён из сокращённой д. н. ф. монотонной булевой функции $f(\tilde{x})$, так как $K_{\tilde{\alpha}}$ является единственным простым импликантом, покрывающим набор $\tilde{\alpha}$.

Проведённые рассуждения показывают, что для монотонной булевой функции проблемы, связанные с минимизацией д. н. ф., в значительной степени отпадают. Её сокращённая д. н. ф. устроена достаточно просто. Она состоит из конъюнкций, соответствующих нижним единицам и является одновременно её кратчайшей и минимальной д. н. ф. Отметим этот важный результат.

Теорема 1. Сокращённая д. н. ф. монотонной булевой функции не содержит отрицаний переменных и является её единственной кратчайшей и минимальной д. н. ф. Максимально возможная длина этой д. н. ф. равна $C_n^{[n/2]}$ и достигается на функции, множество нижних единиц которой совпадает со слоем $[n/2]$.

Чтобы проиллюстрировать, как на практике возникают монотонные булевы функции, обратимся к уже рассматривавшейся в разделе 4 задаче о покрытии строк бинарной таблицы её столбцами, не предполагая теперь, что это именно таблица Куайна. Пусть имеется бинарная таблица размерности $m \times n$, самого общего вида, в каждой строке которой имеется хотя бы одна единица. На 2^n подмножествах множества её столбцов определим булеву функцию следующим образом. Полагаем её равной единице, если подмножество является покрытием строк матрицы, и нулю — в противном случае. Очевидно, что введенная таким образом функция будет монотонной. Нижними единицами такой функции являются неизбыточные покрытия, т. е. такие подмножества столбцов, которые являются покрытиями, но перестают быть таковыми при удалении любого входящего в них столбца.

Рассмотрим теперь вопрос о числе монотонных булевых функций от n переменных. Заметим, прежде всего, что любое подмножество антицепи также является антицепью. Это позволяет, взяв в качестве исходной антицепь мощности $C_n^{[n/2]}$, имеющую $2^{C_n^{[n/2]}}$ подмножеств, сразу получить нижнюю оценку для числа монотонных булевых функций.

Утверждение 2. Имеется не менее $2^{C_n^{[n/2]}}$ монотонных булевых функций от n переменных.

Если брать другие слои куба и рассматривать в качестве нижних единиц их подмножества, то к уже полученным $2^{C_n^{[n/2]}}$ функциям добавится огромное число новых монотонных функций. Однако, как было показано в 1969 году американским учёным Д. Клейтменом (D. Kleitman) [9], несложная нижняя оценка утверждения 1 даёт правильную асимптотику логарифма для числа монотонных булевых функций, т. е. число монотонных булевых функций равно $2^{C_n^{[n/2]}(1+o(1))}$, $n \to \infty$. Позднее российским учёным А. Д. Коршуновым [11] была найдена асимптотика и самого числа монотонных булевых функций. Она описывается двумя различными формулами при чётном и нечётном n и вид её достаточно сложен. Но устроена типичная монотонная булева функция, как было установлено А. Д. Коршуновым, достаточно просто. Уточним, прежде всего, само понятие «типичной функции».

Говорят, что некоторое свойство выполнено для почти всех булевых функций $f(x_1,\ldots,x_n)$ из некоторого класса функций, если доля функций в этом классе, для которых оно не выполняется, стремится к нулю с ростом n.

Оказывается, что почти все монотонные функции от n переменных имеют нижние единицы лишь в трёх средних слоях, а именно в слоях с номерами m, $m-1$ и $m+1$, где $m=n/2$ при n чётном и $m=\lceil n/2\rceil$ или $m=\rceil n/2\lceil$ при n нечётном. При этом в слое m лежит асимптотически $C_n^{n/2}/2$ нижних единиц, а суммарное число нижних единиц в слоях с номерами $m-1$ и $m+1$ не превышает $2^{n/2}$, т. е. составляет ничтожно малую часть от числа нижних единиц в слое m.

Экстремальные свойства шаров Хэмминга. Шары Хэмминга обладают в дискретном метрическом пространстве B^n рядом интересных экстремальных свойств. Одним из таких свойств является их изопериметрическое свойство. Классическая изопериметрическая задача на плоскости состоит в отыскании фигуры, имеющей при заданном периметре наибольшую площадь. Как известно, такой фигурой является круг. В трёхмерном пространстве телом заданного объёма с минимальной поверхностью является сфера. Здесь будет рассмотрен аналог этой задачи для дискретного метрического пространства B^n, где множеством с минимальной границей оказывается шар Хэмминга. Все последующие результаты этого раздела принадлежат советскому математику Р. Г. Нигматуллину (1939–1986).

Пусть $A\subseteq B^n$. Вершина $\tilde{\alpha}\in A$ называется внутренней вершиной множества A, если $S_1(\tilde{\alpha})\in A$ (все соседние с $\tilde{\alpha}$ вершины лежат в A). В противном случае $\tilde{\alpha}$ называется граничной вершиной множества A. Множество $\Gamma(A)$ всех граничных точек множества A называется его границей.

Границей шара $\hat{S}_r(\tilde{\alpha})$ является сфера $S_r(\tilde{\alpha})$, мощность которой равна C_n^r. Оказывается, что это минимально возможная мощность границы для множества мощности $\chi(n,r)$.

Теорема 2 [14]. Для множеств мощности $\chi(n,r)$ минимум мощности границы достигается на шаре радиуса r и только на нём: $\min\limits_{|A|=\chi(n,r)} |\Gamma(A)| = C_n^r$.

Все имеющиеся к настоящему времени доказательства этой теоремы достаточно длинны и технически сложны. Поэтому, опуская доказательство, рассмотрим важные следствия, получаемые из этой теоремы. С оригинальным доказательством Р. Г. Нигматуллина заинтересованный читатель может ознакомиться по первоисточнику [14] или по монографии [15], где имеются ссылки и на работы других авторов.

С изопериметрическим свойством шаров Хэмминга тесно связано их свойство максимизации расстояния между множествами.

Пусть $P,Q \subseteq B^n$. Расстоянием между множествами P и Q называется величина

$$D(P,Q) = \min_{\tilde{\gamma}\in P, \tilde{\delta}\in Q} d(\tilde{\gamma},\tilde{\delta}).$$

Такое определение расстояния между множествами является общепринятым, хотя оно и не является метрикой. Оно симметрично $D(P,Q) = D(Q,P)$, но $D(P,Q)=0$ не значит, что $P=Q$, а означает, что $P\cap Q\neq\varnothing$. Не имеет место также и неравенство треугольника.

Для $1\le p, q\le 2^n-1$ положим $D(p,q) = \max_{|P|=p,\,|Q|=q} D(P,Q)$. Таким образом, $D(p,q)$ обозначает то максимальное расстояние, на которое могут быть разнесены два множества с мощностями p и q. Отметим, что $D(p,q)$ является монотонно невозрастающей функцией по обоим аргументам, т. е. $p'\ge p$ и $q'\ge q$, то $D(p',q')\le D(p,q)$.

Нашей целью является доказательство того, что максимальное расстояние между двумя множествами мощности $\chi(n,k)$ и $\chi(n,l)$, где $l+k\le n$, равно $n-l-k$. Для этого нам потребуются ещё одно определение и ряд вспомогательных утверждений.

Окаймлением $O(P)$ множества $P\subseteq B^n$ называется множество $O(P)=\Gamma\left(B^n\setminus P\right)$, другими словами, $O(P)$ — это множество всех вершин $\tilde{x}\in B^n\setminus P$, для которых $D\left(\{\tilde{x}\},P\right)=1$.

В качестве следствия из теоремы 2 получаем

Утверждение 3. Для множеств мощности $\chi(n,r)$ минимум мощности окаймления достигается на шаре радиуса r и только на нём

$$\min_{|P|=\chi(n,r)} |O(P)| = C_n^{r+1}.$$

Доказательство. Если $|P| = \chi(n,r)$, то $|B^n \setminus P| = \chi(n,n-r-1)$, и утверждение вытекает из теоремы 2. □

При доказательстве последующей теоремы будет использована

Лемма 1. Если $D(P,Q)>0$, то $D\big(P,Q\cup O(Q)\big) = D(P,Q)-1$.

Доказательство. Кратчайшая связывающая P и Q цепь длины $D(P,Q)$, своим последним звеном соединяет некоторую вершину из $Q\cup O(Q)$ с некоторой вершиной из Q. Убрав это звено, получим кратчайшую связывающую P и $Q\cup O(Q)$ цепь длиной $D(P,Q)-1$. □

А теперь получим важный результат, который будет использован в дальнейшем в Дополнении 2 при изучении статистических свойств булевых функций.

Теорема 3 [14].

Если $k+l \le n-1$, то $D\big(\chi(n,k),\chi(n,l)\big) = n-k-l$.

Доказательство. Взяв два шара радиусов k и l с центрами в противоположных вершинах куба, убеждаемся, что

$$D\big(\chi(n,k),\chi(n,l)\big) \ge n-k-l.$$

Доказательство того, что $D\big(\chi(n,k),\chi(n,l)\big) \le n-k-l$, проведём индукцией по параметру $d = n-k-l$. При $d=1$, если множества мощности $\chi(n,k)$ и $\chi(n,l)$ не пересекаются, то они заполняют весь куб B^n, так как

$$\chi(n,k)+\chi(n,l) = \sum_{i=0}^{k} C_n^i + \sum_{i=0}^{n-k-1} C_n^i = \sum_{i=0}^{k} C_n^i + \sum_{i=0}^{n-k-1} C_n^{n-i} =$$

$$= \sum_{i=0}^{k} C_n^i + \sum_{j=k+1}^{n} C_n^j = \sum_{i=0}^{n} C_n^i = 2^n,$$

и расстояние между множествами равно единице.

Пусть утверждение справедливо при $n-k-l = d-1$. Докажем его для $n-k-l = d$, т. е. покажем, что $D\big(\chi(n,k),\chi(n,l-1)\big) = n-k-l+1$. Пусть $|P| = \chi(n,k)$, $|Q| = \chi(n,l-1)$. Покажем, что $D(P,Q) \le n-k-l+1$. Рассмотрим $D\big(P,Q\cup O(Q)\big)$. Согласно утверждению 3 $O(Q) \ge C_n^l$, поэтому $|Q\cup O(Q)| \ge \chi(n,l)$, и по предположению индукции $D\big(P,Q\cup O(Q)\big) \le$ $\le n-k-l$. Но тогда согласно лемме 1 $D(P,Q) \le n-k-l+1$. □

1. Какова мощность интервала, соответствующего в n-мерном кубе элементарной конъюнкции x_1x_2?

   а) 2; б) 4; в) 2^{n-2}.

2. Сколько рёбер имеет n-мерный куб?

   а) 2^n; б) C_n^2; в) $n2^{n-1}$.

3. Чему в n-мерном кубе равна мощноеть пересечения двух сфер Хэмминга единичного радиуса, расстояние между центрами которых также равно единице?

   а) 2; б) n; в) 0.

Ответы: 1 — в, 2 — в, 3 — а.

2.7. Полные системы функций. Теорема Поста

В разделе 4 было показано, что любая булева функция может быть представлена формулой в $\{\vee, \&, \neg\}$, т. е. выражена через дизъюнкцию, конъюнкцию и отрицание. Такое представление может быть получено, например, с помощью дизъюнктивной или конъюнктивной нормальных форм.

Пусть $A=\{f_1, f_2, \ldots\}$ — некоторое множество булевых функций. Замыканием A называется множество $[A]$ всех тех булевых функций, которые могут быть выражены формулами в A.

Так как любая булева функция может быть выражена через дизъюнкцию, конъюнкцию и отрицание, то $[\{\vee, \&, \neg\}] = P_2$.

Множество булевых функций $\{f_1, f_2, \ldots\}$ называется полной системой функций, если $[\{f_1, f_2, \ldots\}] = P_2$.

Таким образом, $\{\vee, \&, \neg\}$ — полная система функций. Из правил де Моргана имеем

$$x_1 \vee x_2 = \overline{\bar{x}_1 \& \bar{x}_2} \text{ и } x_1 \& x_2 = \overline{\bar{x}_1 \vee \bar{x}_2}.$$

Поэтому $\{\&, \neg\}$ и $\{\vee, \neg\}$ — также полные системы. Заметим, однако, что система $\{\&, \vee\}$ не является полной, так как и дизъюнкция, и конъюнкция являются монотонными функциями, а любая суперпозиция монотонных функций, как легко понять, также является монотонной функцией. Таким образом, класс $[\{\&, \vee\}]$ лежит внутри класса монотонных функций. Соображения, подобные этим, будут играть важную роль в дальнейшем.

Полная система функций называется базисом в P_2, если при удалении из неё любой из функций система перестает быть полной.

В полных системах функций {&, ¬} и {∨, ¬} ни одна из функций не может быть удалена. В самом деле, с помощью отрицания могут быть получены только x_i и $\overline{x}_i$, а с помощью операций & и ∨ без использования отрицания могут быть получены лишь монотонные булевы функции. Поэтому системы {&, ¬} и {∨, ¬} являются базисами.

Базисы могут состоять и из одной функции. Примерами таких базисов являются штрих Шеффера и стрелка Пирса. Это следует из равенств

$$\overline{x} = x \mid x = x \downarrow x;\ x_1 \,\&\, x_2 = (x_1 \mid x_2) \mid (x_1 \mid x_2);\ x_1 \vee x_2 = (x_1 \downarrow x_2) \downarrow (x_1 \downarrow x_2).$$

Функция $f(x_1,\ldots,x_n)$ называется шефферовой, если она образует базис в P_2.

Штрих Шеффера и стрелка Пирса являются примерами шефферовых функций от двух переменных.

Интересным базисом является система {1, &, ⊕}. Её полнота следует из того, что $\overline{x} = x \oplus 1$. Если в формуле, записанной в этой системе, воспользовавшись законом дистрибутивности конъюнкции по отношению к сложению по модулю 2: $x_1(x_2 \oplus x_3) = x_1x_2 \oplus x_1x_3$, раскрыть все скобки, то получатся элементарные конъюнкции без отрицаний переменных, складываемые по mod 2, т. е. полином по mod 2. Каждая элементарная конъюнкция в такой сумме может встретиться не более одного раза, так как $K \oplus K = 0$. Такие полиномы называют полиномами Жегалкина, в честь советского математика и логика И. И. Жегалкина (1869–1947), рассмотревшего базис {1, &, ⊕} в 1927 году. Максимальный ранг входящих в полином конъюнкций является степенью полинома.

Константу 0 будем считать пустым полиномом, т. е. полиномом, не содержащим ни одной конъюнкции. При необходимости константу 0 можно представить также как $0 = 1 \oplus 1$.

Теорема 1 (И. И. Жегалкин, 1927). Для каждой булевой функции существует единственный представляющий её полином Жегалкина.

Доказательство. Возможность такого представления следует из полноты системы функций {1, &, ⊕}. Для доказательства единственности достаточно заметить, что всего существуют 2^n элементарных конъюнкций, составленных из n переменных и не содержащих отрицаний пере-

менных, включая тождественно равную единице пустую конъюнкцию. Поэтому существуют 2^{2^n} полиномов Жегалкина — ровно столько же, сколько булевых функций от n переменных. □

Рассмотрим представление булевой функции с помощью полинома Жегалкина на примере дизъюнкции. Зная на основании теоремы 1, что такое представление существует, можно записать

$$x_1 \vee x_2 = a_{12}x_1x_2 \oplus a_1x_1 \oplus a_2x_2 \oplus a_0,$$

где $a_{12}, a_1, a_2, a_0 \in \{0,1\}$ — неизвестные коэффициенты. Подставляя $x_1 = x_2 = 0$, получаем $a_0 = 0$. Подставляя $x_1 = 0, x_2 = 1$, получаем $a_2 = 1$. Подставляя $x_1 = 1, x_2 = 0$, получаем $a_1 = 1$. И, наконец, подставляя $x_1 = 1$, $x_2 = 1$, получаем $1 = a_{12} \oplus 1 \oplus 1$, т. е. $a_{12} = 1$. Таким образом, имеем

$$x_1 \vee x_2 = x_1x_2 \oplus x_1 \oplus x_2.$$

Это представление можно получить и проще, воспользовавшись правилом де Моргана:

$$x_1 \vee x_2 = \overline{\bar{x}_1\bar{x}_2} = (x_1 \oplus 1)(x_2 \oplus 1) \oplus 1 = x_1x_2 \oplus x_1 \oplus x_2.$$

Булева функция $f(x_1,\ldots,x_n)$ называется линейной, если линеен её полином Жегалкина, т. е.

$$f(x_1,\ldots,x_n) = a_0 \oplus a_1x_1 \oplus a_2x_2 \oplus \ldots \oplus a_nx_n.$$

Линейная функция задается с помощью $n+1$ коэффициентов $a_0, a_1, \ldots, a_n \in \{0,1\}$, поэтому существуют 2^{n+1} линейных функций от n переменных. Заметим также, что каждая линейная функция принимает единичное значение ровно на половине наборов значений своих переменных.

Во всех вышерассмотренных случаях полнота некоторой системы функций доказывалась путем получения из этой системы дизъюнкции, конъюнкции и отрицания, образующих полную систему. Если в исследуемую на полноту систему входит небольшое число функций, зависящих лишь от одной или двух переменных, то получение из них дизъюнкции, конъюнкции и отрицания или доказательство невозможности такого получения, как правило, не представляет значительных трудностей. В других случаях, однако, подобное исследование может оказаться непростым. Эффективный критерий полноты был указан американским математиком Постом (E. L. Post, 1897–1954). К изложению этого критерия и связанных с ним результатов мы теперь и переходим.

Множество функций $A = \{f_1, f_2, \ldots\}$ называется замкнутым классом, если его замыкание совпадает с ним самим: $[A] = A$.

Согласно утверждению 5.5, множество всех самодвойственных функций является замкнутым классом. Ясно также, что замкнутыми классами являются множество всех монотонных функций и множество всех линейных функций. Эти три замкнутых класса принято обозначать соответственно S, M и L.

Введём ещё два замкнутых класса:

1) класс $T_0 = \{f(x_1,\ldots,x_n) : f(0,\ldots,0) = 0\}$, который состоит из функций, принимающих на нулевом наборе значение 0, и называется классом функций, сохраняющих 0;
2) класс $T_1 = \{f(x_1,\ldots,x_n) : f(1,\ldots,1) = 1\}$, который состоит из функций, принимающих на единичном наборе значение 1, и называется классом функций, сохраняющих 1.

Утверждение 1. Ни один из замкнутых классов T_0, T_1, S, M, L не лежит внутри никакого другого, т. е. они попарно несравнимы по включению.

Доказательство этого утверждения представлено таблицей 1, в которой символами «+» и «–» показаны соответственно принадлежность и непринадлежность классам T_0, T_1, S, M, L функций 0, 1, $\bar{x}$, x_1x_2, $x_1x_2 \oplus x_1x_2 \oplus x_2x_3$. Как непосредственно видно из таблицы, классы T_0, T_1, S, M, L несравнимы по включению на множестве из этих пяти функций. □

Функции \ Классы	T_0	T_1	S	M	L
0	+	–	–	+	+
1	–	+	–	+	+
$\bar{x}$	–	–	+	–	+
x_1x_2	+	+	–	+	–
$x_1x_2 \oplus x_1x_2 \oplus x_2x_3$	+	+	+	–	–

Таблица 1

Ясно, что если некоторое множество функций A целиком лежит внутри одного из замкнутых классов T_0, T_1, S, M, L, то это множество не может быть полной системой функций, так как все суперпозиции функций из A также будут лежать внутри этого же класса. Таким образом, для того чтобы система была полной, необходимо, чтобы она не лежала целиком ни в одном из классов T_0, T_1, S, M, L (классы Поста). Оказывается, что это условие является также и достаточным.

Теорема 2 (критерий полноты Поста). Для того, чтобы некоторое множество булевых функций было полной системой, необходимо и достаточно, чтобы оно не лежало целиком ни в одном из 5 замкнутых классов T_0, T_1, S, M, L.

Доказательству теоремы предпошлём три леммы, относящиеся к несамодвойственной, немонотонной и нелинейной функциям.

Лемма 1. Если $f(x_1,\ldots,x_n)\notin \mathrm{S}$, то подстановкой x или $\bar{x}$ вместо каждой переменной x_i из неё можно получить несамодвойственную функцию одной переменной, т. е. константу.

Доказательство. Так как $f\notin \mathrm{S}$, то найдётся булев набор $(\sigma_1,\ldots,\sigma_n)$ такой, что $f(\bar{\sigma}_1,\ldots,\bar{\sigma}_n)=f(\sigma_1,\ldots,\sigma_n)$. Рассмотрим функцию $\varphi(x)=f(x^{\sigma_1},\ldots,x^{\sigma_n})$. Имеем

$$\varphi(0)=f(0^{\sigma_1},\ldots,0^{\sigma_n})=f(\bar{\sigma}_1,\ldots,\bar{\sigma}_n)=f(\sigma_1,\ldots,\sigma_n)=f(1^{\sigma_1},\ldots,1^{\sigma_n})=\varphi(1).$$ □

Лемма 2. Если $f(x_1,\ldots,x_n)\notin \mathrm{M}$, то подстановкой 0, 1 или x вместо каждой переменной x_i из неё можно получить функцию $\bar{x}$.

Доказательство. Из того, что $f\notin \mathrm{M}$, следует, что найдутся два булевых набора $\tilde{\alpha}$ и $\tilde{\beta}$ такие, что $\tilde{\alpha}<\tilde{\beta}$ и $f(\tilde{\alpha})=1$, а $f(\tilde{\beta})=0$. Можно считать, что $d(\tilde{\alpha},\tilde{\beta})=1$, так как, если $d(\tilde{\alpha},\tilde{\beta})>1$, то, двигаясь из вершины $\tilde{\alpha}$ в вершину $\tilde{\beta}$ по рёбрам куба, получим в момент переключения функции с 1 на 0 два соседних набора с такими же свойствами. Пусть наборы $\tilde{\alpha}$ и $\tilde{\beta}$ отличаются в i-й координате, т. е.

$$\tilde{\alpha}=(\alpha_1,\ldots,\alpha_{i-1},0,\alpha_{i+1},\ldots,\alpha_n),\ \tilde{\beta}=(\alpha_1,\ldots,\alpha_{i-1},1,\alpha_{i+1},\ldots,\alpha_n).$$

Тогда $\varphi(x)=f(\alpha_1,\ldots,\alpha_{i-1},x,\alpha_{i+1},\ldots,\alpha_n)=\bar{x}$. □

Лемма 3. Если $f(x_1,\ldots,x_n)\notin \mathrm{L}$, то, имея константы и отрицание, из неё можно получить конъюнкцию $x_1 \& x_2$.

Доказательство. В силу нелинейности функции f, в её полиноме Жегалкина найдётся член, являющийся произведением не менее двух переменных. Без ограничения общности можно считать, что в него входят x_1 и x_2. Тогда полином можно привести к виду

$$f(x_1,\ldots,x_n)=x_1x_2f_{12}(x_3,\ldots,x_n)\oplus x_1f_1(x_3,\ldots,x_n)\oplus$$
$$\oplus x_2f_2(x_3,\ldots,x_n)\oplus f_0(x_3,\ldots,x_n),$$

где $f_{12}(x_3,\ldots,x_n)$ не равна тождественно нулю.

Пусть $\alpha_3,\ldots,\alpha_n\in\{0,1\}$ таковы, что $f_{12}(\alpha_3,\ldots,\alpha_n)=1$. Тогда

$$\varphi(x_1,x_2)=f(x_1,x_2,\alpha_3,\ldots,\alpha_n)=x_1x_2\oplus\beta x_1\oplus\gamma x_2\oplus\delta,$$

где $\beta,\gamma,\delta\in\{0,1\}$. Тогда требуемой конъюнкцией является функция

$$\varphi(x_1,x_2)=\varphi(x_1\oplus\gamma,\ x_2\oplus\beta)\ \oplus\ \beta\gamma\oplus\delta=x_1x_2.$$ □

Доказательство теоремы Поста. Необходимость уже была показана. Докажем достаточность. Из того, что система функций не лежит целиком ни в одном из 5 замкнутых классов, следует, что в ней можно выделить подсистему из не более чем пяти функций $\{f_{\overline{T}_0}, f_{\overline{T}_1}, f_{\overline{S}}, f_{\overline{M}}, f_{\overline{L}}\}$, не принадлежащих соответственно классам T_0, T_1, S, M, L (среди функций $f_{\overline{T}_0}, f_{\overline{T}_1}, f_{\overline{S}}, f_{\overline{M}}, f_{\overline{L}}$ могут быть и одинаковые). Докажем, что множество функций $\{f_{\overline{T}_0}, f_{\overline{T}_1}, f_{\overline{S}}, f_{\overline{M}}, f_{\overline{L}}\}$ образует полную систему.

Покажем сначала, как, используя только функции $f_{\overline{T}_0}, f_{\overline{T}_1}, f_{\overline{S}}$, можно получить константы 1 и 0. Если $f_{\overline{T}_0}(1,\ldots,1)=1$ и $f_{\overline{T}_1}(0,\ldots,0)=0$, то функции $\varphi_0(x)=f_{\overline{T}_0}(x,\ldots,x)$ и $\varphi_1(x)=f_{\overline{T}_1}(x,\ldots,x)$ являются соответственно константами 1 и 0. А если хотя бы одно из этих равенств не имеет места, например первое, то $\varphi_0(x)=f_{\overline{T}_0}(x,\ldots,x)=\overline{x}$. Тогда с помощью функции $f_{\overline{S}}$, воспользовавшись леммой 1, получим одну из констант, а взяв её отрицание — и другую.

Получив константы 1 и 0, с помощью функции $f_{\overline{M}}$, опираясь на лемму 2, получаем $\overline{x}$. И, наконец, имея 0, 1 и $\overline{x}$, с помощью функции $f_{\overline{L}}$, опираясь на лемму 3, получаем x_1x_2. Это даёт полную систему $\{\&,\ \neg\}$, что и доказывает теорему. □

В качестве примера использования теоремы для исследования полноты системы функций исследуем на полноту систему из двух функций $\{x_1\to x_2,\ x_1\to\overline{x}_2x_3\}$. Составим таблицу принадлежности этих функций пяти замкнутым классам (табл. 2).

Функции \ Классы	T_0	T_1	S	M	L
$x_1\to x_2$	–	+	–	–	–
$x_1\to\overline{x}_2x_3$	–	–	–	–	–

Таблица 2

Быстро проверить правильность заполнения таблицы можно следующим образом. Принадлежность или непринадлежность классам T_0 и T_1 проверяется непосредственно. Непринадлежность классам обеих функций классам S и L следует из того, что любая функция в каждом из этих классов ровно на половине наборов обязана обращаться в 1, а на другой полови-

не — в 0. Представив каждую из функций в виде д. н. ф.: $x_1 \to x_2 = \bar{x}_1 \vee x_2$ и $x_1 \to \bar{x}_2 x_3 = \bar{x}_1 \vee \bar{x}_2 x_3$, легко убедиться в том, что единичных наборов у каждой из функций больше. Немонотонность функции $x_1 \to x_2$ следует из сравнения её значений на наборах (00) и (10), а немонотонность функции $x_1 \to \bar{x}_2 x_3$ — из сравнения её значений на наборах (000) и (100).

Так как в каждом столбце таблицу имеется хотя бы один минус, данная система функций является полной. Но она не является базисом, так как одна функция $x_1 \to \bar{x}_2 x_3$ уже образует полную систему и, следовательно, базис. Посмотрим, как из неё можно получить, например, отрицание и конъюнкцию. Обозначим для удобства нашу функцию как $f(x_1, x_2, x_3)$:

$$f(x_1, x_2, x_3) = x_1 \to \bar{x}_2 x_3 = \bar{x}_1 \vee \bar{x}_2 x_3 .$$

Подставив в $f(x_1, x_2, x_3)$ вместо всех переменных x_1, получаем:

$$f(x_1, x_1, x_1) = \bar{x}_1 \vee \bar{x}_1 x_1 = \bar{x}_1 \vee 0 = \bar{x}_1 .$$

Таким образом, отрицание получено. Получим теперь константу 1, воспользовавшись леммой 1. Так как $f(1,0,1) = f(0,1,0)$, имеем:

$$f(x_1, \bar{x}_1, x_1) = \bar{x}_1 \vee x_1 x_1 = 1 .$$

Теперь, подставив в $f(x_1, x_2, x_3) = \bar{x}_1 \vee \bar{x}_2 x_3$ константу 1 вместо x_1, $\bar{x}_2$ — вместо x_2, x_1 — вместо x_3, получаем конъюнкцию $x_1 x_2$:

$$x_1 x_2 = f(1, \bar{x}_2, x_1) = f\big(f(x_1, \bar{x}_1, x_1), f(x_2, x_2, x_2), x_1\big) =$$

$$= f\big(f(x_1, f(x_1, x_1, x_1), x_1), f(x_2, x_2, x_2), x_1\big).$$

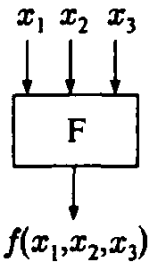


Рис. 1

Полученную для конъюнкции формулу можно наглядно представить с помощью схемы. Пусть функция $f(x_1, x_2, x_3) = x_1 \to \bar{x}_2 x_3$ реализуется функциональным элементом F (рис. 1).

Тогда схема на рис. 2 реализует конъюнкцию $x_1 x_2$.

Как непосредственно следует из теоремы Поста, базис в P_2 не может состоять более чем из 5 функций. Эта верхняя граница может быть уточнена.

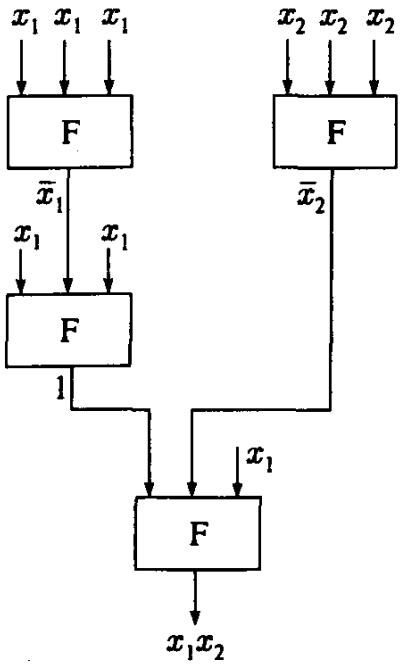


Рис. 2

Утверждение 2. Базис в P_2 может состоять самое большее из четырёх функций.

Доказательство. При доказательстве теоремы 2 полная система функций $\{f_{\overline{T}_0}, f_{\overline{T}_1}, f_{\overline{S}}, f_{\overline{M}}, f_{\overline{L}}\}$ состояла из не более чем пяти различных функций. Как было видно из доказательства, функция $f_{\overline{T}_0}$ либо не является самодвойственной, либо не сохраняет 1 и не является монотонной. В первом случае полной является система $\{f_{\overline{T}_0}, f_{\overline{T}_1}, f_{\overline{M}}, f_{\overline{L}}\}$, во втором — система $\{f_{\overline{T}_0}, f_{\overline{S}}, f_{\overline{L}}\}$, т. е. в обоих случаях четырёх функций оказывается достаточно.

Покажем, что существует базис из четырёх функций. Рассмотрим систему функций $\{0, 1, x_1x_2, x_1 \oplus x_2 \oplus x_3\}$. Как видно из табл. 3, эта система является полной (в каждом столбце имеется хотя бы один минус), а удаление из неё любой функции ведёт к тому, что она утрачивает свойство полноты. Поэтому система представляет собой базис из четырёх функций. □

Функции \ Классы	T_0	T_1	S	M	L
0	+	−	−	+	+
1	−	+	−	+	+
x_1x_2	+	+	−	+	−
$x_1 \oplus x_2 \oplus x_3$	+	+	+	−	+

Таблица 3

В качестве непосредственного следствия из теоремы Поста получаем

Следствие 1. Любой замкнутый класс функций в P_2, не совпадающий с P_2, лежит внутри одного из кассов T_0, T_1, S, M, L.

Замкнутый класс A называется предполным классом, если $A \neq P_2$, а для любой функции $f \notin A$ имеет место $[A \cup f] = P_2$.

Можно сказать, что предполный класс — это максимальный по включению замкнутый класс, не совпадающий с P_2.

Теорема 3. В P_2 предполными являются классы T_0, T_1, S, M, L и только они.

Доказательство. Если к любому из 5 замкнутых классов T_0, T_1, S, M, L добавить не принадлежащую ему функцию, то вследствие утверждения 1 полученная система не будет целиком лежать ни в одном из

классов T_0, T_1, S, M, L и по теореме Поста её замыкание совпадёт с P_2. Поэтому T_0, T_1, S, M, L являются предполными классами.

Пусть K — некоторый предполный класс. Если бы K не лежал целиком ни в одном из 5 замкнутых классов, то по теореме Поста он совпал бы с P_2, что невозможно. Он не может быть также и собственным подмножеством ни одного из замкнутых классов T_0, T_1, S, M, L, так как в этом случае существовала бы функция f такая, что класс $[K \cup \{f\}]$ продолжал бы лежать внутри того же самого замкнутого класса. Поэтому K должен совпасть с одним из классов T_0, T_1, S, M, L. □

Понятие базиса, как полной и неизбыточной системы функций, может применяться не только к P_2, но и к любому замкнутому классу в P_2. Например, в классе M монотонных функций базис, как нетрудно убедиться, образуют функции $\{0,1,\vee,\&\}$. В связи с этим отметим, что, как было показано Постом, множество замкнутых классов в P_2 счётно и каждый из них имеет конечный базис (см.[20]).

Вопросы для самопроверки

1. К каким из 5 замкнутых классов T_0, T_1, S, M, L принадлежит функция $x_1x_2 \vee x_2x_3 \vee x_1x_3$?

   а) T_0, T_1, M; б) T_0, T_1, L; в) T_0, T_1, S, M.

2. Выразить функцию $x_1 \vee \overline{x}_2$ в базисе $\{1, \&, \oplus\}$.

   а) $x_1 \oplus x_2$; б) $x_1 \oplus x_2 \oplus x_1x_2$ в) $x_1x_2 \oplus x_2 \oplus 1$.

3. Какие из трёх систем функций 1) $\{\rightarrow, \neg\}$, 2) $\{\sim, \oplus\}$, 3) $\{\&, \rightarrow, \oplus\}$ являются базисами в P_2?

   а) 1; б) 1, 2 и 3; в) 1 и 2.

Ответы: 1 — в, 2 — в, 3 — а.

2.8. Пороговая логика

Булевы функции и голосующие процедуры. Пусть коллективом из n членов принимается решение, состоящее в выборе одной из двух возможных альтернатив, причём каждый член коллектива высказывается в пользу одной из альтернатив, а коллективное решение принимается на основе индивидуальных предпочтений, т. е. решение принимается с помощью некоторой голосующей процедуры. Обозначая альтернативы символами 0 и 1, получаем, что коллективное решение является некоторой булевой функцией $f: \{0,1\}^n \rightarrow \{0,1\}$.

Если n — нечётное число, то наиболее распространенным способом коллективного решения является решение по большинству, т. е. выбор той

из альтернатив, которая собрала больше голосов. Этот способ задаётся *мажоритарной* булевой функцией, которая равна 0 или 1 в зависимости от того, каких символов больше в двоичном наборе. Поучительно проанализировать методами булевой алгебры этот издревле используемый человечеством метод коллективного решения и обосновать выбор мажоритарной функции, рассмотрев ряд естественных свойств, выполнение которых желательно для решающей функции $f(x_1,\ldots,x_n)$.

Во-первых, равноправие членов коллектива ведёт к тому, что булева функция не должна изменяться при произвольной перестановке своих аргументов, т. е. обязана быть симметрической.

Во-вторых, равноправие альтернатив означает, что $f(\overline{x}_1,\ldots,\overline{x}_n)=\overline{f}(x_1,\ldots,x_n)$, т. е. функция должна быть самодвойственной.

В-третьих, функция не должна уменьшаться от замены некоторых из нулей набора на единицы, т. е. должна быть монотонной.

Теорема 1. При нечётном n мажоритарная функция — это единственная булева функция, являющаяся одновременно симметрической, самодвойственной и монотонной. При чётном n подобной функции не существует.

Доказательство. Из условия симметричности следует, что $f(x_1,\ldots,x_n)$ постоянна на каждом слое единичного n-мерного куба, так как подстановкой переменных любой набор слоя может быть переведён в любой другой набор этого же слоя. Из самодвойственности вытекает, что $n=2k+1$, так как при $n=2k$ условия симметричности и самодвойственности несовместимы на k-м слое. В самом деле, отрицание переменных переводит каждый набор k-го слоя снова в набор этого же слоя. Согласно условию симметричности значения функции на двух наборах должны совпадать, а согласно условию самодвойственности — быть противоположными. И, наконец, из монотонности и самодвойственности следует, что $f(x_1,\ldots,x_n)=0$ на всех слоях от нулевого до k-го и $f(x_1,\ldots,x_n)=1$ на всех слоях от $(k+1)$-го до $(2k+1)$-го, т. е. функция является мажоритарной. □

Пороговые функции. Хотя д. н. ф. и является универсальным способом представления булевых функций, она весьма неудобна для представления мажоритарной функции. При $n=2k+1$ нижние единицы мажоритарной функции занимают $(k+1)$-й слой целиком и длина её д. н. ф. оказывается равной C_{2k+1}^{k+1}, что представляет собой слишком большое число, чтобы такое представление могло быть реально выписано. Мажоритарную функцию удобно задать следующим образом. Выпишем неравенство

$$x_1+\ldots+x_n\le n/2,$$

и будем считать, что $f(x_1,\ldots,x_n)=0$, если набор $(x_1,\ldots,x_n)$ удовлетворяет неравенству, и $f(x_1,\ldots,x_n)=1$ в противном случае.

Булева функция $f:\{0,1\}^n \to \{0,1\}$ называется пороговой, если существует линейное неравенство с действительными коэффициентами

$$a_1x_1+\ldots+a_nx_n \le b, \qquad (1)$$

которое выполнено на тех и только тех наборах $\tilde{x}=(x_1,\ldots,x_n)$, для которых $f(\tilde{x})=0$. Коэффициенты a_i называются весами, $(a_1,\ldots,a_n)$ — весовым вектором, b — порогом.

Мажоритарная функция является, таким образом, пороговой. Общую пороговую функцию можно рассматривать как булеву функцию, реализующую процедуру взвешенного голосования, когда значимость отдельных членов коллектива выражается приписанными им весами. Изучение возможностей представления булевых функций с помощью неравенства (1) представляет собой самостоятельный раздел, называемый *пороговой логикой*. Он богат теоретическими результатами различного характера и имеет прикладное значение.

Впервые пороговые функции были рассмотрены в 1943 году американскими учёными Уорреном Маккаллоком (W. S. McCulloch) и Вальтером Питтсом (W. Pitts) при построении ими математической модели функционирования нервной клетки — нейрона. Продолжая их исследования, американский учёный Фрэнк Розенблатт (Frank Rosenblatt (1928–1971)) предложил в 1957 году адаптивный алгоритм обучения порогового элемента, привлёкший внимание специалистов по искусственному интеллекту. Пороговые функции продолжают изучаться и использоваться в системах искусственного интеллекта, называемых нейронными сетями. Они также естественным образом возникают в задачах дискретной оптимизации, которые будут рассмотрены в Главе 4. В этом разделе будут рассмотрены различные аспекты пороговой логики.

При изучении пороговых функций геометрическая интерпретация с помощью единичного n-мерного куба играет особенно важную роль. Пороговая функция от двух переменных $f(x_1,x_2)$ в плоскости $<x_1,x_2>$ задаётся прямой: $a_1x_1+a_2x_2=b$, отделяющей единичные вершины функции от нулевых. (См. рис. 1, где $f(x_1,x_2)=x_1x_2$.)

Пороговая функция от трёх переменных $f(x_1,x_2,x_3)$ задаётся плоскостью в пространстве $<x_1,x_2,x_3>$: $a_1x_1+a_2x_2+a_3x_3=b$, рассекающей единичный куб так, что в вершинах по одну сторону плоскости функция равна нулю, а по другую — единице. На рис. 2 плоскость задаёт функцию $f(x_1,x_2,x_3)=x_1x_2 \vee x_2x_3$.

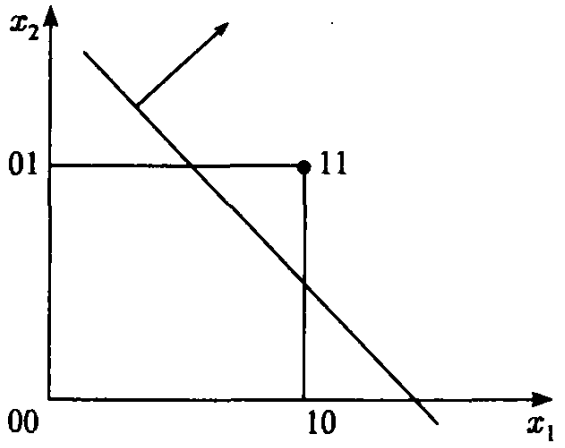


Рис. 1

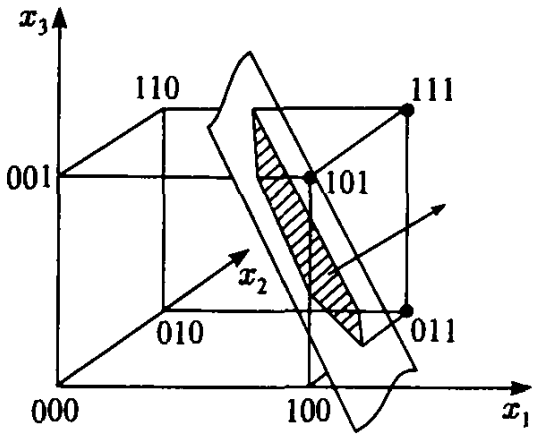


Рис. 2

В общем случае n переменных можно считать, что в евклидовом пространстве $< x_1,\ldots,x_n >$ пороговая функция задаётся гиперплоскостью $a_1x_1 + \ldots + a_nx_n = b$, рассекающей единичный n-мерный куб так, что в вершинах по одну сторону гиперплоскости функция равна нулю, а по другую — единице. При этом, как легко понять, небольшим изменением порога всегда можно добиться строгой разделимости, т. е. строгого выполнения неравенства (1) на нулевых вершинах.

Обозначим через N_n число пороговых функций от n переменных. Все четыре булевы функции от одной переменной являются, очевидно, пороговыми, а из 16 булевых функций от двух переменных, как непосредственно видно из геометрической интерпретации, не являются пороговыми лишь две: $x_1 \oplus x_2$ и $x_1 \oplus x_2 \oplus 1$ — счётчик чётности и его отрицание. Следовательно, имеем $N_1 = 4$, $N_2 = 14$.

Для перечисления пороговых функций от трёх переменных полезно рассмотреть классы пороговых функций с заданным числом единиц. Пусть $N_n(M)$ — число пороговых функций от n переменных, принимающих значение 1 на M наборах. $N_n(M)$ можно также считать числом пороговых множеств мощности M, понимая под *пороговым множеством* подмножество вершин единичного n-мерного куба, которое может быть отделено от своего дополнения гиперплоскостью. С каждым пороговой функцией связаны два пороговых множества — множество её единиц и множество её нулей. Если каждой пороговой функции поставить в соответствие пороговое множество её единиц, то соответствие будет взаимно однозначным.

Замечая, что $N_n(M) = N_n(2^n - M)$, получаем

$$N_3 = 2(N_3(0) + N_3(1) + N_3(2) + N_3(3)) + N_3(4).$$

Имеем $N_3(0) = 1$, $N_3(1) = 8$ (множество вершин куба), $N_3(2) = 12$ (множество рёбер куба). Пороговое множество мощности 3 состоит из

трёх вершин, лежащих в одной грани. Таких множеств в каждой грани имеется 4, поэтому $N_3(3) = 6 \cdot 4 = 24$.

Пороговых множеств мощности 4, как можно непосредственно геометрически убедиться, имеется всего два типа: 1) четыре вершины, лежащие в одной грани; 2) вершина с тремя соседними вершинами. Поэтому $N_3(4) = 6 + 8 = 14$, а полное число пороговых функций от трёх переменных $N_3 = 104$.

Подобное исследование числа пороговых функций в пространствах большего числа измерений уже не может быть выполнено столь наглядными геометрическими методами и требует более глубоких исследований. Однако уже на этих элементарных примерах заметна тенденция к уменьшению доли пороговых функций с ростом n.

Здесь естественным образом возникает вопрос о критерии, позволяющем определять, является ли таблично заданная булева функция от n переменных пороговой. Часто оказывается удобным простое условие, называемое *2-суммируемостью*, которое достаточно для того, чтобы функция не была пороговой.

Булева функция f называется 2-суммируемой, если существуют наборы $\tilde{\alpha}, \tilde{\beta} \in f^{-1}(0)$ и наборы $\tilde{\gamma}, \tilde{\delta} \in f^{-1}(1)$ такие, что $\tilde{\alpha} + \tilde{\beta} = \tilde{\gamma} + \tilde{\delta}$, где под суммой понимается обычное покоординатное сложение векторов в евклидовом пространстве.

Утверждение 1. Если функция 2-суммируема, то она не является пороговой.

Доказательство. Допустим противное, что функция пороговая, т. е. существует неравенство (1), отделяющее единичные наборы от нулевых. Тогда $\begin{cases} a_1\alpha_1 + \ldots + a_n\alpha_n \le b \\ a_1\beta_1 + \ldots + a_n\beta_n \le b \end{cases}$ и $\begin{cases} a_1\gamma_1 + \ldots + a_2\gamma_2 > b \\ a_1\delta_1 + \ldots + a_2\delta_2 > b \end{cases}$. Складывая неравенства в первой и во второй системе, получаем

$$a_1(\alpha_1 + \beta_1) + \ldots + a_n(\alpha_n + \beta_n) \le b \text{ и } a_1(\gamma_1 + \delta_1) + \ldots + a_n(\gamma_n + \delta_n) > b,$$

что невозможно ввиду $\tilde{\alpha} + \tilde{\beta} = \tilde{\gamma} + \tilde{\delta}$. □

Непороговость функции $x_1 \oplus x_2$ сразу вытекает из её 2-суммируемости: $(0,1) + (1,0) = (1,1) = (0,0) + (1,1)$.

Аналогично 2-суммируемости можно определить *k-суммируемость* булевой функции как равенство суммы некоторых её k нулевых наборов сумме k единичных наборов, где в каждой из сумм могут быть повто-

ряющиеся слагаемые. Если для некоторого k функция k-суммируема, то, как легко аналогично предыдущему показать, она не является пороговой. Кроме того, из общей теоремы выпуклого анализа, согласно которой множества $f^{-1}(0)$ и $f^{-1}(1)$ линейно разделимы в том и только в том случае, если их выпуклые оболочки не пересекаются, следует, что, если функция не является k-суммируемой ни при каком натуральном k, то она пороговая. Такую полную несуммируемость можно считать критерием пороговости, однако этот критерий трудно проверяем и поэтому не играет существенной роли в исследовании пороговых функций.

Заметим, что, если в неравенстве (1) все веса a_i неотрицательны, то задаваемая им пороговая функция является, очевидно, монотонной. Верно и более сильное утверждение.

Утверждение 2. Пороговая функция, задаваемая неравенством (1) с положительными весами, является монотонной, а монотонная пороговая функция всегда может быть задана неравенством (1) с неотрицательными весами.

Доказательство. В доказательстве нуждается лишь вторая часть утверждения теоремы. Допустим, что неравенство

$$-a_1x_1 + a_2x_2 + \ldots + a_nx_n \le b,$$

где $a_1 \ge 0$, задаёт монотонную пороговую функцию $f(\tilde{x})$. Покажем, что переменная x_1 в таком случае является несущественной и неравенство

$$a_2x_2 + \ldots + a_nx_n \le b \qquad (2)$$

задаёт ту же самую функцию $f(\tilde{x})$. Пусть неравенство (2) задаёт функцию $f'(\tilde{x})$. В подкубе $x_1 = 0$ функции $f(\tilde{x})$ и $f'(\tilde{x})$, очевидно, совпадают. Допустим, что они не совпадают в подкубе $x_1 = 1$. Тогда найдётся набор $\tilde{\alpha} = (\alpha_2, \ldots, \alpha_n)$ такой, что $f(1,\tilde{\alpha}) = 0$, а $f'(1,\tilde{\alpha}) = 1$. Имеем

$$f(0,\tilde{\alpha}) = f'(0,\tilde{\alpha}) = f'(1,\tilde{\alpha}) = 1.$$

Таким образом, $f(1,\tilde{\alpha}) = 0$, а $f(0,\tilde{\alpha}) = 1$, что противоречит условию монотонности. □

Булева функция называется однородной, если заменой некоторых из переменных на их отрицания из неё может быть получена монотонная функция.

Замена переменной x_1 на её отрицание в пороговой логике эквивалентна замене в неравенстве (1) переменной x_1 на $1 - x_1$. Вновь полученная функция будет задаваться неравенством

$$-a_1x_1 + \ldots + a_nx_n \le b - a_1,$$

т. е. коэффициент при x_1 сменит знак. Подобными заменами можно добиться, чтобы все коэффициенты a_i в задающем пороговую функцию неравенстве стали неотрицательными, преобразовав, таким образом, исходную пороговую функцию в монотонную. Отсюда вытекает

Утверждение 3. Пороговая функция является однородной.

В качестве следствия из утверждения 2 и теоремы 6.1 получаем

Утверждение 4. Сокращённая д. н. ф. пороговой функции содержит каждую существенную переменную либо только с отрицанием, либо только без отрицания и является её единственной кратчайшей и минимальной д. н. ф. Максимально возможная длина этой д. н. ф. равна $C_n^{[n/2]}$.

Простейшими пороговыми множествами являются подкубы — интервалы, соответствующие элементарным конъюнкциям.

Утверждение 5. Элементарная конъюнкция и её отрицание являются пороговыми функциями.

Доказательство. Принимая во внимание сделанные выше замечания относительно преобразования неравенства (1) при замене переменных на их отрицания, утверждение теоремы достаточно доказать для конъюнкций, не содержащих отрицаний переменных. Пусть элементарная конъюнкция имеет вид $x_{i_1} x_{i_2} \ldots x_{i_r}$. Тогда уравнение

$$x_{i_1} + x_{i_2} + \ldots + x_{i_r} = r - 1/2$$

задаёт гиперплоскость, отделяющую единичные вершины конъюнкции от нулевых. □

Таким образом, длина д. н. ф. пороговой функции может варьироваться от 1 до $C_n^{[n/2]}$. Что можно сказать о длине д. н. ф. в типичном случае? Как было показано О. В. Шабаниным [19], почти все пороговые функции имеют экспоненциально большую длину д. н. ф..

Вариация порога. Покажем теперь, что пороговых функций значительно больше, чем элементарных конъюнкций. Заметим, прежде всего, что, если в задающем пороговую функцию неравенстве (1) непрерывно изменять значение порога b от $-\infty$ до $+\infty$, то гиперплоскость будет перемещаться параллельно самой себе так, что число нулевых вершин функции будет увеличиваться от 0 до 2^n. Будем считать, что при таком перемещении гиперплоскости в любой момент в ней находится не более одной

вершины гиперкуба. В противном случае этого можно добиться небольшим изменением весов $a_1, \ldots, a_n$, не меняющим исходной функции. Поэтому при вариации порога возникнет $2^n + 1$ различных пороговых функций.

Сделав это замечание, проследим, как изменяется число пороговых функций при увеличении числа переменных на единицу.

Утверждение 6. Для числа N_i пороговых функций от i переменных справедливо следующее рекуррентное соотношение: $N_{i+1} \geq (2^i + 1)N_i$.

Доказательство. Сопоставим каждой пороговой функции от i переменных, задаваемой линейным неравенством $a_1x_1 + \ldots + a_ix_i \leq b$, множество пороговых функций от $i+1$ переменных, задаваемых неравенствами вида $a_1x_1 + \ldots + a_ix_i + ax_{i+1} \leq b$, где a пробегает всё множество действительных значений от $-\infty$ до $+\infty$. При этом в подкубе $x_{i+1} = 0$ варьирование a не меняет функции, а в подкубе $x_{i+1} = 1$ гиперплоскость $a_1x_1 + \ldots + a_ix_i = b - a$, перемещаясь параллельно самой себе и пересекая 2^i вершин, порождает $2^i + 1$ различных функций. Тем самым каждой функции от i переменных сопоставлено множество из $2^i + 1$ функций от $i+1$ переменных и эти N_i множеств попарно не пересекаются. □

Полученное рекуррентное соотношение позволяет получить нижнюю оценку для числа пороговых функций.

Следствие 1. Для числа N_n пороговых функций от n переменных справедлива оценка

$$N_n \geq \prod_{i=1}^{n-1}(2^i + 1) > \prod_{i=1}^{n-1} 2^i = 2^{n(n-1)/2} = 2^{(1-o(1))n^2/2}. \qquad (3)$$

Чтобы понять, насколько оценка (3) близка к истинному числу пороговых функций, необходима верхняя оценка для их числа.

Параметры Чоу (Chow's parameters). Введём важное для пороговой логики понятие параметров Чоу булевой функции. Сложив как векторы в евклидовом пространстве все единичные для функции $f(\tilde{x})$ наборы, получим целочисленный n-мерный вектор

$$(s_1, \ldots, s_n) = \sum_{\tilde{x} \in f^{-1}(1)} \tilde{x}.$$

Дополнив его нулевой координатой $s_0 = |f^{-1}(1)|$, равной числу единичных вершин функции, получим $(n+1)$-мерный вектор $s(f) = (s_0, s_1, \ldots, s_n)$, который называется *вектором параметров Чоу* булевой функции $f(\tilde{x})$.

Теорема 2. Пусть $f(\tilde{x})$ — пороговая функция. Если $g(\tilde{x})$ — некоторая булева функция (не обязательно пороговая) такая, что $s(g) = s(f)$, то $g(\tilde{x}) = f(\tilde{x})$.

Доказательство. Пусть функция $f(\tilde{x})$ задаётся неравенством (1). По условию теоремы имеем

$$\sum_{\tilde{x} \in f^{-1}(1)} \tilde{x} = \sum_{\tilde{x} \in g^{-1}(1)} \tilde{x}.$$

Выбросив из обеих сумм общую часть, получаем

$$\sum_{\tilde{x} \in f^{-1}(1) \setminus g^{-1}(1)} \tilde{x} = \sum_{\tilde{x} \in g^{-1}(1) \setminus f^{-1}(1)} \tilde{x}. \qquad (4)$$

Условие равенства нулевых компонент векторов Чоу даёт

$$|f^{-1}(1) \setminus g^{-1}(1)| = |g^{-1}(1) \setminus f^{-1}(1)| = m.$$

Скалярно помножив обе части равенства (4) на весовой вектор $(a_1, \ldots, a_n)$, на основании неравенства (1) заключаем, что при $m \neq 0$ левая часть полученного равенства будет строго больше, чем bm, а правая не превышает этой величины. Полученное противоречие показывает, что $m = 0$ и $f(\tilde{x}) = g(\tilde{x})$. □

По вектору Чоу произвольной булевой функции можно, в принципе, установить, является ли она пороговой, а если является, то и однозначно её идентифицировать. Поэтому векторы Чоу могут быть использованы для табулирования пороговых функций. Каждая компонента вектора Чоу принимает целочисленное значение в диапазоне от 0 до 2^n. Отсюда следует, что число различных векторов Чоу заведомо не превышает $(2^n+1)^{n+1}$. Это даёт верхнюю оценку для числа пороговых функций.

Следствие 2. Для числа N_n пороговых функций от n переменных справедлива оценка

$$N_n \leq (2^n+1)^{n+1} = 2^{n^2(1+o(1))}. \qquad (5)$$

Пороговые представления. На вопрос о том, какая из оценок (3) или (5) точнее, ответ будет дан несколько позже. В любом случае пороговые функции составляют исчезающе малую часть от множества всех булевых функций. Однако любая булева функция может быть задана системой линейных неравенств.

Теорема 3. Каждая булева функция $f(x_1,\ldots,x_n)$ может быть задана системой линейных неравенств

$$\begin{aligned} a_{11}x_1+\ldots+a_{1n}x_n &\le b_1,\\ &\ldots\ldots\ldots\ldots\\ a_{m1}x_1+\ldots+a_{mn}x_n &\le b_m \end{aligned} \qquad (6)$$

так, что нулевые для функции наборы удовлетворяют системе, а единичные — не удовлетворяют. Такое задание булевой функции называется её пороговым представлением, а наименьшее число необходимых для этого неравенств — пороговым числом $t(f)$ булевой функции f.

Доказательство. Пороговое представление произвольной булевой функции можно получить, опираясь на утверждение 5 и заменив в её д. н. ф. каждую элементарную конъюнкцию соответствующим линейным неравенством. □

Пороговое представление можно рассматривать как дизъюнкцию пороговых функций. Геометрически оно сводится к отсечению гиперплоскостями всех единичных вершин функции f. Так как каждому неравенству системы (6) соответствует пороговое множество, то пороговое представление можно также рассматривать как покрытие единичных вершин булевой функции f пороговыми множествами. Подобно длине кратчайшей д. н. ф. пороговое число является мерой сложности булевой функции.

Если представление с помощью д. н. ф. удобно для булевых функций с малым числом единиц, а с помощью к. н. ф. — с малым числом нулей, то пороговое представление сочетает в себе, как было показано Л. И. Липкиным, оба этих достоинства. □

Теорема 4 [12]. Для порогового числа произвольной булевой функции f справедливо неравенство

$$t(f) \le \min\{|f^{-1}(1)|, |f^{-1}(0)|\}.$$

Доказательство. Соотношение $t(f) \le |f^{-1}(1)|$ очевидно, так как каждую единичную вершину можно отсечь одним линейным неравенством. Докажем соотношение $t(f) \le |f^{-1}(0)|$ индукцией по числу нулевых вершин. При $|f^{-1}(0)|=1$ оно верно. Пусть оно верно для всех булевых функций с числом нулей, не превышающим $k-1$, и пусть функция $f(x_1,\ldots,x_n)$ имеет $k \ge 2$ нулей. Тогда существует переменная x_j такая, что в каждом из подкубов $x_j=0$ и $x_j=1$ имеются нулевые вершины функции. Не теряя общности, будем считать, что $j=n$. Пусть в подкубе $x_n=1$ лежит m нулей функции, а в подкубе $x_n=0$ — $k-m$ нулей, где

$m \le k-1$ и $k-m \le k-1$. Тогда по предположению индукции существует система из m неравенств

$$a_{11}x_1 + \ldots + a_{1n-1}x_{n-1} \le b_1,$$
$$\ldots\ldots\ldots\ldots$$
$$a_{m1}x_1 + \ldots + a_{mn-1}x_{n-1} \le b_m,$$

задающая функцию f в подкубе $x_n = 1$, и система из $k-m$ неравенств

$$a_{m+11}x_1 + \ldots + a_{m+1n-1}x_{n-1} \le b_{m+1},$$
$$\ldots\ldots\ldots\ldots$$
$$a_{k1}x_1 + \ldots + a_{kn-1}x_{n-1} \le b_k,$$

задающая функцию f в подкубе $x_n = 0$.

Тогда система из k неравенств

$$a_{11}x_1 + \ldots + a_{1n-1}x_{n-1} + M x_n \le b_1 + M,$$
$$\ldots\ldots\ldots\ldots$$
$$a_{m1}x_1 + \ldots + a_{mn-1}x_{n-1} + Mx_n \le b_m + M,$$
$$a_{m+11}x_1 + \ldots + a_{m+1n-1}x_{n-1} - Mx_n \le b_{m+1},$$
$$\ldots\ldots\ldots\ldots$$
$$a_{k1}x_1 + \ldots + a_{kn-1}x_{n-1} - Mx_n \le b_k,$$

где M достаточно велико, задаёт функцию f. Первые m неравенств отсекают все единичные вершины функции f в подкубе $x_n = 1$, не затрагивая вершин подкуба $x_n = 0$, а последние $k-m$ неравенств отсекают все единичные вершины функции f в подкубе $x_n = 0$, не затрагивая вершин подкуба $x_n = 1$. □

Следствие 3. Максимальное значение порогового числа у булевых функций от n переменных, равное 2^{n-1}, достигается ровно на двух функциях: счётчике чётности и его отрицании.

Доказательство. Счётчик чётности и его отрицание являются, как нетрудно убедиться, единственными булевыми функциями, у которых $|f^{-1}(1)| = |f^{-1}(0)| = 2^{n-1}$, а множество $f^{-1}(1)$ состоит из изолированных вершин. □

Однородное представление. Во многих задачах пороговой логики в качестве булевой функции f вместо отображения $f\colon \{0,1\}^n \to \{0,1\}$ удобнее рассматривать отображение $f\colon \{-1,1\}^n \to \{-1,1\}$. В геометрической

интерпретации это соответствует переходу от гиперкуба $[0, 1]^n$ к гиперкубу $[-1, 1]^n$, центр которого совпадает с началом координат. Введя обозначение $\tilde{y} = (y_1, \ldots, y_n) \in \{-1, 1\}^n$, пороговую функцию (1) можно записать в виде

$$f(\tilde{y}) = \operatorname{sgn}(a_0 + a_1 y_1 + \ldots + a_n y_n), \text{ где } \operatorname{sgn}(z) = \begin{cases} 1, & z > 0 \\ -1, & z \le 0. \end{cases} \qquad (7)$$

При этом, как обычно, можно считать, что $(a_0 + a_1 y_1 + \ldots + a_n y_n) \neq 0$ при $\tilde{y} \in \{-1, 1\}^n$.

Связь между переменными x_i и y_i задаётся соотношением $y_i = 2x_i - 1$, т. е. значению $x_i = 0$ соответствует значение $y_i = -1$, а значению $x_i = 1$ — значение $y_i = 1$. Коэффициенты $a_1, \ldots, a_n$ здесь те же, что и в (1), а *обобщённый порог*, или *смещение* $a_0 = \sum_{i=1}^{n} a_i - 2b$. Теперь операция отрицания сводится к постановке перед отрицаемым выражением знака «–». Функция, двойственная (7), запишется как

$$f^*(\tilde{y}) = -\operatorname{sgn}(a_0 - a_1 y_1 - \ldots - a_n y_n) = \operatorname{sgn}(-a_0 + a_1 y_1 + \ldots + a_n y_n). \qquad (8)$$

Как видно из сравнения (7) и (8), при $a_0 = 0$ функция (7) является самодвойственной.

Задание пороговой функции в форме (7) можно сделать ещё изящнее, если, дополнив набор переменных нулевой координатой, значение которой постоянно и равно единице, рассмотреть вектор $\mathbf{y} = (1, y_1, \ldots, y_n)$, а также ввести *расширенный вектор весов* $\mathbf{a} = (a_0, a_1, \ldots, a_n)$, сделав смещение его нулевой координатой. Тогда, используя скалярное произведение, (7) можно переписать как

$$f(\mathbf{y}) = \operatorname{sgn}(\mathbf{a}, \mathbf{y}). \qquad (9)$$

При таком представлении, которое принято называть *однородным*, смещение a_0 входит в весовой вектор $\mathbf{a}$ равноправно с весами, а исходную булеву функцию от n переменных $(y_1, \ldots, y_n)$ можно рассматривать как подфункцию в подкубе $y_0 = 1$ самодвойственной функции от $n+1$ переменных $(y_0, y_1, \ldots, y_n)$. Этим, в частности, устанавливается взаимно однозначное соответствие между пороговыми функциями от n переменных и самодвойственными пороговыми функциями от $n+1$ переменных. Отметим этот факт.

Утверждение 7. Число пороговых функций от n переменных равно числу самодвойственных пороговых функций от $n+1$ переменных.

Нахождение весовых коэффициентов. Рассмотрим теперь следующий важный для пороговой логики вопрос. Пусть пороговая булева функция задана таблично. Как найти реализующий её весовой вектор? Эта задача сводится к нахождению решения $(a_0, a_1, \ldots, a_n)$ системы из 2^n неравенств вида

$$a_0 + a_1 y_1 + \ldots + a_n y_n < 0$$

или

$$a_0 + a_1 y_1 + \ldots + a_n y_n > 0,$$

где для каждого набора $\tilde{y} \in \{-1,1\}^n$ знак неравенства выбирается в зависимости от значения функции $f(\tilde{y})$ на этом наборе, и может решаться стандартными средствами линейного программирования. Здесь, однако, возможно и иное, весьма простое решение, основное на однородном представлении (9).

Выбрав произвольно некоторый начальный расширенный вектор весов, будем циклически обходить вершины гиперкуба. В случае правильного знака скалярного произведения в данной вершине переходим к следующей вершине, не изменяя весового вектора. В случае же ошибки прибавляем к расширенному вектору весов или вычитаем из него расширенный вектор-вершину в зависимости от того, какое из действий приводит к изменению скалярного произведения в нужном направлении. Таким образом, на $(k+1)$-м шаге при просмотре вершины $\mathbf{y}_k$ весовой вектор корректируется по правилу:

$$\mathbf{a}_{k+1} = \begin{cases} \mathbf{a}_k, & \operatorname{sgn}(\mathbf{a}_k, \mathbf{y}_k) = f(\mathbf{y}_k) \\ \mathbf{a}_k + f(\mathbf{y}_k)\mathbf{y}_k, & \operatorname{sgn}(\mathbf{a}_k, \mathbf{y}_k) \neq f(\mathbf{y}_k). \end{cases}$$

Данный алгоритм, предложенный в 1957 году американским учёным Фрэнком Розенблаттом (1928–1971), впервые продемонстрировал адаптивные возможности порогового элемента. Названный своим создателем «персептроном», он породил целое направление в искусственном интеллекте и распознавании образов — теорию искусственных нейронных сетей.

Теорема 5. За конечное число шагов указанный алгоритм коррекции весов приводит к весовому вектору, реализующему пороговую функцию f.

Доказательство. Так как функция f является пороговой, существует расширенный вектор весов $\mathbf{a}$ такой, что $(\mathbf{a},\mathbf{y}) > 0$, если $f(\mathbf{y}) = 1$, и $(\mathbf{a},\mathbf{y}) < 0$, если $f(\mathbf{y}) = -1$, для всех $\mathbf{y}$. Выберем некоторое $\Delta > 0$. Умножение вектора $\mathbf{a}$ на произвольную положительную константу не изменяет

пороговой функции. Поэтому всегда можно считать, что для всех $\mathbf{y}$ выполнено неравенство $|(\mathbf{a},\mathbf{y})| > \dfrac{\Delta+n+1}{2}$.

Рассмотрим квадрат разности $(\mathbf{a}_k - \mathbf{a})^2$ между весовым вектором $\mathbf{a}_k$, полученным после k шагов алгоритма, и вектором $\mathbf{a}$, реализующим пороговую функцию f, и оценим изменение этой величины на $(k+1)$-м шаге в том случае, если на этом шаге весовой вектор был изменён, т. е. $\operatorname{sgn}(\mathbf{a}_k,\mathbf{y}_k) \neq f(\mathbf{y}_k)$ и $\mathbf{a}_{k+1} = \mathbf{a}_k + f(\mathbf{y}_k)\mathbf{y}_k$:

$$(\mathbf{a}_{k+1}-\mathbf{a})^2-(\mathbf{a}_k-\mathbf{a})^2=\left(\mathbf{a}_k+f(\mathbf{y}_k)\mathbf{y}_k-\mathbf{a}\right)^2-(\mathbf{a}_k-\mathbf{a})^2=$$

$$=\mathbf{y}_k{}^2+2f(\mathbf{y}_k)(\mathbf{a}_k,\mathbf{y}_k)-2f(\mathbf{y}_k)(\mathbf{a},\mathbf{y}_k)<-\Delta,$$

так как $\mathbf{y}_k^2 = n+1$, $f(\mathbf{y}_k)(\mathbf{a}_k,\mathbf{y}_k) < 0$, $f(\mathbf{y}_k)(\mathbf{a},\mathbf{y}_k) > \dfrac{\Delta+n+1}{2}$.

Таким образом, при каждом изменении весового вектора рассматриваемая величина уменьшается не меньше, чем на Δ. А так как эта величина неотрицательна, то в процессе работы алгоритма произойдёт лишь конечное число изменений весового вектора, после чего будет получен вектор, безошибочно реализующий заданную пороговую функцию f. □

Разбиение пространства гиперплоскостями. Рассмотрим теперь вопрос о том, что представляет в $(n+1)$-мерном евклидовом пространстве $< a_0, a_1, \ldots, a_n >$ множество весовых векторов $\mathbf{a} = (a_0, a_1, \ldots, a_n)$, задающих с помощью (9) некоторую фиксированную пороговую функцию $f(\mathbf{y})$. Обратимся вначале к хорошо известным фактам аналитической геометрии. На плоскости $< x_1, x_2 >$ линейной формой $L(x_1,x_2) = a_1x_1 + a_2x_2$ задаётся прямая $L(x_1,x_2) = 0$, проходящая через начало координат и рассекающая плоскость на две *открытых полуплоскости*: $L(x_1,x_2) < 0$ и $L(x_1,x_2) > 0$. Каждая из полуплоскостей является *открытым множеством*, так как вместе с каждой точкой содержит и некоторую её окрестность. Если взять m таких прямых, то они рассекут плоскость на $2m$ областей.

Рассматривая линейную форму как скалярное произведение $L(x_1,x_2) = (\mathbf{a},\mathbf{x})$, где $\mathbf{a} = (a_1,a_2)$, $\mathbf{x} = (x_1,x_2)$, получаем, что прямая $(\mathbf{a}\ \mathbf{x}) = 0$ является множеством векторов $\mathbf{x}$, ортогональных (перпендикулярных) вектору $\mathbf{a}$. Вектор $\mathbf{a}$ называется *нормалью* к прямой, а множество векторов $\mathbf{x}$ образует *одномерное векторное пространство*.

В трёхмерном пространстве $< x_1, x_2, x_3 >$ линейная форма $L(x_1,x_2,x_3) = a_1x_1 + a_2x_2 + a_3x_3 = (\mathbf{a},\mathbf{x})$ задаёт плоскость $(\mathbf{a}, \mathbf{x}) = 0$, проходящую

через начало координат и рассекающую пространство на два *открытых полупространства*: $(\mathbf{a},\mathbf{x})<0$ и $(\mathbf{a},\mathbf{x})>0$. Плоскость H: $(\mathbf{a},\mathbf{x})=0$ является *двумерным векторным пространством*, ортогональным нормали $\mathbf{a}$.

Если имеется m плоскостей, проходящих через начало координат, то они рассекают пространство на множество открытых многогранных конусов с общей вершиной в начале координат. Задачу об определении числа этих конусов впервые рассмотрел в 1826 году знаменитый швейцарский геометр Якоб Штейнер (1796–1863).

Число конусов зависит, вообще говоря, от взаимного расположения плоскостей. Максимальное число их возникает в том случае, когда система плоскостей находится в общем положении, т. е. единственной общей точкой каждых трёх плоскостей является начало координат. В этом случае число конусов зависит только от числа плоскостей.

Теорема 6 (Я. Штейнер, 1826). Максимальное число открытых многогранных конусов, возникающих при рассечении пространства m плоскостями, проходящими через начало координат, равно

$$2\sum_{i=0}^{2} C_{m-1}^{i},$$

и этот максимум достигается в том и только в том случае, если плоскости находятся в общем положении.

Доказательство. Доказательство проведём индукцией по числу плоскостей. Одна плоскость делит пространство на 2 части, что находится в соответствии с формулой. Пусть формула справедлива для m плоскостей. Докажем её для $m+1$ плоскостей. Пусть имеется $m+1$ плоскостей $H_1,\ldots,H_m,H_{m+1}$. По предположению индукции число конусов, на которые разбивают пространство m плоскостей $H_1,\ldots,H_m$, не превышает $2\sum_{i=0}^{2} C_{m-1}^{i}$ и равно этому числу в том и только в том случае, если плоскости находятся в общем положении. При проведении плоскости H_{m+1}, число конусов увеличивается за счёт того, что некоторые из конусов рассекаются плоскостью H_{m+1} на два конуса. Если обратиться к плоскости H_{m+1}, то каждое такое рассечение соответствует в ней одной из областей, на которые она рассекается прямыми, образованными её пересечениями с другими плоскостями. Поэтому прирост числа конусов равен числу областей, на которые H_{m+1} рассекается этими прямыми. Если плоскости $H_1,\ldots,H_m,H_{m+1}$ находятся в общем положении, то плоскость H_{m+1} рассекается m пря-

мыми, вырезающими в ней $2m$ областей. Таким образом, в случае общего положения $m+1$ плоскостей число конусов будет равно

$$2\sum_{i=0}^{2} C_{m-1}^{i} + 2m = 2\sum_{i=0}^{2} C_{m}^{i},$$

в полном соответствии с утверждением теоремы. Если же плоскости находятся не в общем положении, то число конусов уменьшается. □

Подобно этому линейная форма $L(x_1,\ldots,x_2) = a_1x_1 + \ldots + a_2x_2 = (\mathbf{a}\ \mathbf{x})$ задаёт в n-мерном пространстве $< x_1, \ldots, x_n >$ гиперплоскость $H:(\mathbf{a}\ \mathbf{x})=0$, проходящую через начало координат и разрезающую пространство на два открытых полупространства: $(\mathbf{a}\ \mathbf{x}) < 0$ и $(\mathbf{a}\ \mathbf{x}) > 0$. Гиперплоскость $(\mathbf{a}\ \mathbf{x}) = 0$ является векторным подпространством размерности $n-1$, ортогональным нормали $\mathbf{a}$. Семейство из m таких гиперплоскостей разрезают n-мерное пространство на множество открытых многогранных конусов, каждый из которых характеризуется определённым набором знаков линейных форм. *Гиперплоскости находятся в общем центрированном положении, если пересечение любых i плоскостей при $i \le n$ имеет размерность $n-i$, т. е. их нормали линейно независимы.* Швейцарским математиком Людвигом Шлефли (1814–1895), одним из создателей многомерной геометрии, формула Штейнера была в середине XIX века обобщена на n-мерное пространство.

> **Теорема 6bis (Л. Шлефли).** Максимальное число открытых многогранных конусов, возникающих при рассечении n-мерного евклидова пространства m гиперплоскостями, проходящими через начало координат, равно $2\sum_{i=0}^{n-1} C_{m-1}^{i}$, и этот максимум достигается в том и только в том случае, если гиперплоскости находятся в общем положении.

Доказательство. Доказательство практически без изменения повторяет доказательство для трёхмерного пространства. При добавлении $(m+1)$-й гиперплоскости число конусов возрастает на число областей, на которые эта гиперплоскость рассекается остальными гиперплоскостями. Для гиперплоскости, имеющей размерность $n-1$, это число по предположению индукции в случае общего расположения гиперплоскостей равно $2\sum_{i=0}^{n-2} C_{m-1}^{i}$, а m гиперплоскостей рассекают пространство на $2\sum_{i=0}^{n-1} C_{m-1}^{i}$ конусов. Поэтому полное число конусов равно

$$2\sum_{i=0}^{n-1} C_{m-1}^{i} + 2\sum_{i=0}^{n-2} C_{m-1}^{i} = 2\sum_{i=0}^{n-1} C_{m}^{i} . \qquad \square$$

Покажем теперь, как вышеприведённые результаты могут быть использованы для оценки числа пороговых функций. Для фиксированного набора $(y_1, \ldots, y_n) \in \{-1, 1\}^n$ линейная форма $L(a_0, a_1, \ldots, a_n) = a_0 + a_1 y_1 + \ldots + a_n y_n = (\mathbf{a}, \mathbf{y})$ задаёт в $(n+1)$-мерном пространстве $< a_0, a_1, \ldots, a_n >$ гиперплоскость $H : (\mathbf{a}, \mathbf{y}) = 0$, проходящую через начало координат и рассекающую пространство на два открытых полупространства, так что весовые векторы, взятые из одного полупространства, придают пороговой функции $f(\mathbf{y}) = \operatorname{sgn}(\mathbf{a}, \mathbf{y})$ на наборе $\mathbf{y}$ значение -1, а из другого — значение 1. Рассмотрим теперь множество таких гиперплоскостей для всех $(y_1, \ldots, y_n) \in \{-1, 1\}^n$, т. е. 2^n гиперплоскостей вида

$$a_0 \pm a_1 \pm \ldots \pm a_n = 0 . \qquad (10)$$

Они рассекают пространство $< a_0, a_1, \ldots, a_n >$ на некоторое число многогранных конусов. Если два весовых вектора $\mathbf{a}_1$ и $\mathbf{a}_2$ принадлежат одному и тому же конусу, то, как следует из вышеприведённых рассуждений, $\operatorname{sgn}(\mathbf{a}_1, \mathbf{y}) = \operatorname{sgn}(\mathbf{a}_2, \mathbf{y})$ на всех 2^n наборах $\mathbf{y} = (1, y_1, \ldots, y_n)$, т. е. весовые векторы задают одну и ту же пороговую функцию. И наоборот, если два весовых вектора $\mathbf{a}_1$ и $\mathbf{a}_2$ взяты из разных конусов, то множество наборов $\mathbf{y}$, для которых $\operatorname{sgn}(\mathbf{a}_1, \mathbf{y}) \neq \operatorname{sgn}(\mathbf{a}_2, \mathbf{y})$, не пусто. Таким образом, существует взаимно однозначное соответствие между конусами и пороговыми функциями. Но число конусов можно оценить сверху, пользуясь теоремой Штейнера—Шлефли, и эта оценка будет верхней оценкой для числа пороговых функций:

$$N_n \leq 2\sum_{i=0}^{n} C_{2^n-1}^{i} . \qquad (11)$$

При $n = 1, 2$ гиперплоскости (10) находятся, как нетрудно проверить, в общем положении, и в (11) имеет место знак равенства. Но уже при $n = 3$ гиперплоскости (10) находятся не в общем положении. В этом можно убедиться, рассмотрев, например, следующую матрицу, строки которой составлены из нормалей системы (10):

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix} .$$

Она имеет, как легко проверить, ранг 3, а не 4, необходимый для общего положения. Подобная вырожденность приводит к тому, что число пороговых функций от трёх переменных оказывается равным 104, а не $2\sum_{i=0}^{3} C_7^i = 128$. В силу соотношений

$$2\sum_{i=0}^{n} C_{2^n-1}^i \sim 2\frac{2^{n^2}}{n!} = 2^{n^2(1-o(1))} \qquad (12)$$

верхняя оценка (11) с точностью до асимптотики логарифма совпадает с верхней оценкой (5), и между ними и нижней оценкой (3) остаётся прежний разрыв.

Прежде чем перейти к дальнейшему изложению, которое будет вестись на геометрическом языке, уточним связи между геометрическими и алгебраическими понятиями.

Рассмотрим систему из m линейных однородных уравнений

$$\begin{cases} a_{11}x_1 + \ldots + a_{1n}x_n = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + \ldots + a_{mn}x_n = 0, \end{cases}$$

геометрическим образом каждого из которых является проходящая через начало координат гиперплоскость, задаваемая соответствующей нормалью. Множество её решений, геометрически описываемое как пересечение гиперплоскостей, является векторным пространством размерности $n-r$, где r — *ранг системы — максимальное число линейно независимых уравнений*, или, другими словами, — *максимальное число линейно независимых нормалей*. Ввиду неравенства $r \le n$, определённая таким образом размерность неотрицательна. Равенство её нулю означает, что множество решений состоит из единственной точки — начала координат.

Пусть множество гиперплоскостей $\mathrm{H} = \{H_1, \ldots, H_m\}$. Для каждого из 2^m его подмножеств $\mathrm{H}' \subseteq \mathrm{H}$ через $r(\mathrm{H}')$ будем обозначать ранг соответствующего множества нормалей. Геометрически ранг подмножества определяется как $r(\mathrm{H}') = n - \dim\mathrm{H}'$, где $\dim\mathrm{H}' = \dim\bigcap_{H_i \subseteq \mathrm{H}'} H_i$ — размерность пересечения подмножества гиперплоскостей H'.

Американским математиком Робертом Уиндером, интенсивно работавшим в области пороговой логики в шестидесятые годы прошлого века, была найдена красивая формула для числа конусов, справедливая для произвольного расположения гиперплоскостей, проходящих через начало координат.

Теорема 7 [24]. Число конусов, на которые n-мерное евклидово пространство рассекается множеством центрированных гиперплоскостей $\mathrm{H} = \{H_1, \ldots, H_m\}$, равно

$\sum_{H' \subseteq H} (-1)^{|H'|-r(H')}$, где сумма берётся по всем 2^m подмножествам множества H.

Доказательство. Доказательство проведём индукцией по числу гиперплоскостей. Отметим, прежде всего, что сумма $\sum_{H' \subseteq H} (-1)^{|H'|-r(H')}$ не изменится, если H считать мультимножеством, допуская кратные, т. е. совпадающие гиперплоскости. Согласно (1.2.6) имеем $\sum_{i=1}^{k} (-1)^i C_k^i = -1$ для любого $k = 1, 2, \ldots$. Поэтому рассмотрение вместо одной гиперплоскости k её экземпляров не изменяет ни числа конусов, ни значения суммы.

Далее заметим, что $|H'| - r(H') = \dim H' - (n - |H'|)$. Поэтому

$$\sum_{H' \subseteq H} (-1)^{|H'|-r(H')} = \sum_{H' \subseteq H} (-1)^{\dim H' - (n-|H'|)}.$$

Для одной гиперплоскости утверждение, очевидно, справедливо. Пусть оно справедливо для любых m гиперплоскостей. Обозначим, следуя Уиндеру, число конусов, на которые разбивают n-мерное пространство гиперплоскости $H_1, \ldots, H_m, H_{m+1}$, через $\#\{H_1, \ldots, H_m, H_{m+1}\}^n$. Аналогично доказательству теоремы 6bis имеем

$$\#\{H_1, \ldots, H_m, H_{m+1}\}^n = \#\{H_1, \ldots, H_m\}^n + \#\{H_1 H_{m+1}, \ldots, H_m H_{m+1}\}^{n-1},$$

где $H_i H_{m+1}$ обозначает $(n-2)$-мерную гиперплоскость в $(n-1)$-мерном пространстве H_{m+1}, образованную пересечением H_i с H_{m+1}, а второе слагаемое в правой части равно числу конусов, на которые гиперплоскость H_{m+1} разбивается остальными гиперплоскостями. По индуктивному предположению

$$\#\{H_1, \ldots, H_m\}^n = \sum_{H' \subseteq \{H_1, \ldots, H_m\}} (-1)^{\dim H' - (n-|H'|)},$$

$$\#\{H_1 H_{m+1}, \ldots, H_m H_{m+1}\}^{n-1} = \sum_{H'' \subseteq \{H_1 H_{m+1}, \ldots, H_m H_{m+1}\}} (-1)^{\dim H'' - ((n-1)-|H''|)} =$$

$$= \sum_{\substack{H' \subseteq \{H_1, \ldots, H_m, H_{m+1}\} \\ H_{m+1} \in H'}} (-1)^{\dim H' - (n-|H'|)}.$$

При этом возможные совпадения некоторых из гиперплоскостей $H_i H_{m+1}$, $i = 1, 2, \ldots, m$ в силу сделанного выше замечания не имеют значения. Поэтому

$$\#\{H_1,\ldots,H_m,H_{m+1}\}^n = \sum_{\mathrm{H}'\subseteq\{H_1,\ldots,H_m,H_{m+1}\}} (-1)^{\dim \mathrm{H}'-(n-|\mathrm{H}'|)} =$$

$$= \sum_{\mathrm{H}'\subseteq\{H_1,\ldots,H_m,H_{m+1}\}} (-1)^{|\mathrm{H}'|-r(\mathrm{H}')}. \qquad \square$$

Заметим, что из теоремы Уиндера легко получить формулу Штейнера—Шлефли. Если m гиперплоскостей находятся в n-мерном пространстве в общем положении, то $r(\mathrm{H}')=|H'|$ при $|H'|\le n$ и $r(\mathrm{H}')=n$, при $|H'|\ge n$. Поэтому согласно теореме Уиндера число конусов равно

$$\sum_{i=0}^{n} C_m^i + \sum_{i=n+1}^{m} (-1)^{i-n} C_m^i = \sum_{i=0}^{n} C_m^i - (C_{m-1}^n + C_{m-1}^{n+1}) + (C_{m-1}^{n+1} + C_{m-1}^{n+2}) - \ldots + (-1)^{m-n} =$$

$$= \sum_{i=0}^{n} C_m^i - C_{m-1}^n = (1 + (C_{m-1}^0 + C_{m-1}^1) + (C_{m-1}^1 + C_{m-1}^2) +$$

$$+ (C_{m-1}^2 + C_{m-1}^3) + \ldots + (C_{m-1}^{n-1} + C_{m-1}^n)) - C_{m-1}^n = 2\sum_{i=0}^{n-1} C_{m-1}^i,$$

в полном соответствии с теоремой Штейнера—Шлефли.

С помощью теоремы Уиндера можно оценить число конусов снизу в случае, если расположение гиперплоскостей не является общим. Для системы гиперплоскостей H обозначим через П(H) число линейно зависимых подмножеств гиперплоскостей мощности не более n, т. е. таких подмножеств $\mathrm{H}'\subseteq \mathrm{H}$, что $|\mathrm{H}'|\le n$ и $r(\mathrm{H}')<|\mathrm{H}'|$.

Следствие 4. Число конусов, на которые n-мерное евклидово пространство рассекается множеством центрированных гиперплоскостей $\mathrm{H}=\{H_1,\ldots,H_m\}$, не меньше чем $2\sum_{i=0}^{n-1} C_{m-1}^i - 2\Pi(\mathrm{H})$.

Доказательство. Согласно теореме 7 число конусов есть

$$\sum_{\mathrm{H}'\subseteq \mathrm{H}} (-1)^{|\mathrm{H}'|-r(\mathrm{H}')} = \sum_{\substack{\mathrm{H}'\subseteq \mathrm{H}\\ |\mathrm{H}'|\le n}} (-1)^{|\mathrm{H}'|-r(\mathrm{H}')} + \sum_{\substack{\mathrm{H}'\subseteq \mathrm{H}\\ |\mathrm{H}'|> n}} (-1)^{|\mathrm{H}'|-r(\mathrm{H}')} \ge$$

$$\ge \sum_{i=0}^{n} C_m^i - 2\Pi(\mathrm{H}) + \sum_{\substack{\mathrm{H}'\subseteq \mathrm{H}\\ |\mathrm{H}'|> n}} (-1)^{|\mathrm{H}'|-r(\mathrm{H}')}. \qquad (13)$$

Индукцией по числу гиперплоскостей покажем, что всегда имеет место неравенство

$$\sum_{\substack{H'\subseteq H\\ |H'|>n}} (-1)^{|H'|-r(H')} \geq -C_{m-1}^{n}.$$

Предположив, что неравенство справедливо для любых m гиперплоскостей $H_1, \ldots, H_m$, для $m+1$ гиперплоскостей $H_1, \ldots, H_m, H_{m+1}$ имеем

$$\sum_{\substack{H'\subseteq\{H_1,\ldots,H_m,H_{m+1}\}\\ |H'|>n}} (-1)^{|H'|-r(H')} = \sum_{\substack{H'\subseteq H_1,\ldots,H_m\}\\ |H'|>n}} (-1)^{|H'|-r(H')} +$$

$$+ \sum_{\substack{H'\subseteq\{H_1,\ldots,H_m,H_{m+1}\}\\ H_{m+1}\in H',\ |H'|>n}} (-1)^{|H'|-r(H')}.$$

Для первой суммы по индуктивному предположению

$$\sum_{\substack{H'\subseteq H_1,\ldots,H_m\}\\ |H'|>n}} (-1)^{|H'|-r(H')} \geq -C_{m-1}^{n}.$$

Для второй суммы, рассматривая гиперплоскость H_{m+1} как пространство размерности $n-1$, рассекаемое m гиперплоскостями $H_1H_{m+1}, \ldots, H_mH_{m+1}$, также в силу индуктивного предположения имеем

$$\sum_{\substack{H'\subseteq\{H_1,\ldots,H_m,H_{m+1}\}\\ H_{m+1}\in H',\ |H'|>n}} (-1)^{|H'|-r(H')} \geq -C_{m-1}^{n-1}.$$

Отсюда получаем

$$\sum_{\substack{H'\subseteq\{H_1,\ldots,H_m,H_{m+1}\}\\ |H'|>n}} (-1)^{|H'|-r(H')} \geq -C_{m-1}^{n} - C_{m-1}^{n-1} = -C_{m}^{n}.$$

Теперь с помощью (13) получаем

$$\sum_{H'\subseteq H} (-1)^{|H'|-r(H')} \geq \sum_{i=0}^{n} C_m^i - 2\Pi(H) - C_{m-1}^{n} = 2\sum_{i=0}^{n-1} C_{m-1}^{i} - 2\Pi(H). \qquad \square$$

Оценка, даваемая следствием 4, может оказаться полезной, когда размерность пространства $n\to\infty$, число гиперплоскостей $m=O(n)$, а доля линейно зависимых подмножеств гиперплоскостей стремится к нулю. В этом случае число конусов асимптотически совпадает с их числом при общем расположении гиперплоскостей (см. задачу 10 в Дополнении 2). В то же время данная оценка оказывается бесполезной при подсчёте числа пороговых функций, где $m=2^{n-1}$. Далее будет представлена другая нижняя оценка, позволившая продвинуться в этой задаче.

Освободимся от допущения, что все гиперплоскости проходят через одну точку — начало координат, и будем рассматривать в n-мерном пространстве гиперплоскости общего вида, задаваемые уравнением:

$$a_1 x_1 + \ldots + a_n x_n = b\,. \qquad (14)$$

Множество его решений X может быть представлено в виде

$$X = X_0 + \mathbf{x}_b\,,$$

где $\mathbf{x}_b = (x_1, \ldots, x_n)$ — произвольное решение уравнения (14), а X_0 — множество решений соответствующего однородного уравнения, являющееся векторным подпространством размерности $(n-1)$. Таким образом, множество X можно рассматривать как сдвиг $(n-1)$-мерного векторного пространства X_0 на вектор $\mathbf{x}_b$. Множество X называют $(n-1)$-мерным *аффинным подпространством*, для которого мы сохраним название гиперплоскости и обозначение H.

Множество решений системы линейных неоднородных уравнений

$$\begin{cases} a_{11} x_1 + \ldots + a_{1n} x_n = b_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1} x_1 + \ldots + a_{mn} x_n = b_m, \end{cases}$$

являющееся пересечением m подобных гиперплоскостей также является аффинным подпространством. Оно представимо как некоторое частное решение системы плюс общее решение соответствующей однородной системы и поэтому может рассматриваться как сдвиг векторного пространства решений однородной системы на вектор этого частного решения. Размерностью аффинного подпространства, если оно не пусто, называется размерность соответствующего векторного подпространства решений однородной системы. В частном случае одной гиперплоскости размерность равна $n-1$. Удобно также считать, что любая отрицательная размерность выражает тот факт, что соответствующее аффинное подпространство является пустым множеством.

Понятие общего расположения гиперплоскостей является широко известным и общепринятым.

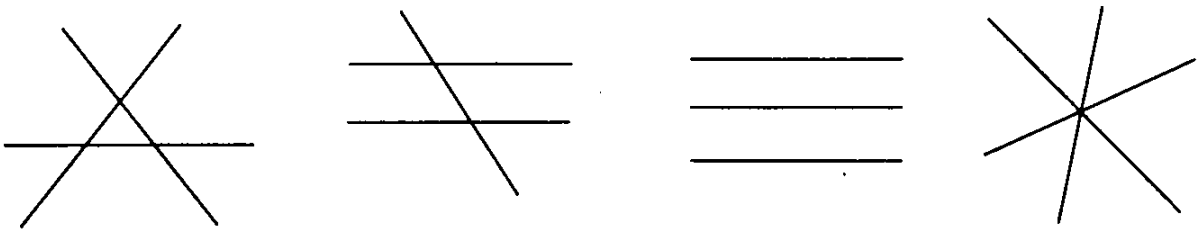
Расположение гиперплоскостей в n-мерном пространстве называется общим (нецентрированным), если пересечение любых i гиперплоскостей имеет размерность $n-i$, при $i > n$ оно пусто.

Определим теперь специальное понятие квазиобщего расположения.

Расположение гиперплоскостей в n-мерном пространстве называется квазиобщим, если пересечение любых i гиперплоскостей имеет размерность $n-i$ или пусто.

Таким образом, при квазиобщем расположении допускается параллельность, но через каждое непустое подпространство пересечения размерности j $(0 \le j \le n-1)$ проходит ровно $n-j$ гиперплоскостей. Общее расположение является, разумеется, квазиобщим.

Эти определения проиллюстрированы на рис. 3, где представлены все возможные случаи расположения трёх прямых на плоскости (трёх гиперплоскостей в двумерном пространстве).



а) *б*) *в*) *г*)

Рис. 3

В случае *а*) расположение общее, в случаях *а*), *б*), *в*) — квазиобщее, случай *г*) не является ни общим, ни квазиобщим, так как здесь через пространство нулевой размерности — точку проходят три прямые.

Сформулируем теперь основной результат.

> **Теорема 8 [6]**. Число открытых областей, на которые евклидово n-мерное пространство рассекается конечным числом гиперплоскостей, всегда не меньше полного числа различных аффинных подпространств, порождённых пересечениями этих гиперплоскостей, считая подпространства всех размерностей от 0 (точки) до $n-1$ (сами гиперплоскости) и n (всё n-мерное пространство). При этом случай равенства имеет место тогда и только тогда, когда расположение гиперплоскостей квазиобщее.

Обращаясь снова к рис. 4, видим, что в случаях *а*), *б*), *в*), когда расположение прямых квазиобщее, число областей действительно совпадает с числом подпространств. Так, в случае *а*) имеется 7 областей, и полное число подпространств пересечения также равно 7: три точки, три прямые и вся плоскость. В случае же *г*), когда расположение не является квазиобщим, имеется 6 областей и только 5 подпространств пересечения: точка, три прямые и вся плоскость.

Доказательство. Доказательство проведём индукцией по числу гиперплоскостей. Одна гиперплоскость разрезает пространство на две области и число подпространств пересечения также равно двум: гиперплоскость и всё пространство. Пусть утверждение теоремы справедливо для евклидовых пространств всех размерностей и m гиперплоскостей. Докажем его для $m+1$ гиперплоскостей. Пусть в n -мерном евклидовом пространстве заданы $m+1$ гиперплоскостей $H_1,\ldots,H_m,H_{m+1}$. По предположению индукции при разрезании пространства гиперплоскостями $H_1,\ldots,H_m$ число

подпространств пересечения не превосходит числа областей. Посмотрим, как изменятся число подпространств и число областей при проведении гиперплоскости H_{m+1}.

Прирост числа областей n-мерного пространства равен числу $(n-1)$-мерных кусков гиперплоскости H_{m+1}, вырезаемых в ней гиперплоскостями $H_1,\ldots,H_m$, так как каждый такой кусок является «перегородкой», делящей одну из прежних областей на две области. По предположению индукции полное число подпространств пересечения в гиперплоскости H_{m+1} не превосходит числа её $(n-1)$-мерных кусков. А так как все вновь появившиеся подпространства пересечения n-мерного пространства лежат в гиперплоскости H_{m+1}, прирост числа подпространств не превышает прироста числа областей и между ними сохраняется прежнее соотношение.

Условие равенства следует из того, что при квазиобщем расположении все подпространства пересечения, принадлежащие H_{m+1}, появляются при проведении гиперплоскости H_{m+1} и прирост числа подпространств равен приросту числа областей. □

При общем положении гиперплоскостей в n-мерном пространстве подпространства, порождённые пересечениями каждых i гиперплоскостей, будут иметь размерность $n-i$ и при $0 \le i \le n$ будут различными подпространствами. В этом случае теорема 8 сразу даёт число областей, впервые найденное другим методом американским математиком Баком, образно сравнившим рассечение пространства с разрезанием на куски большой круглой головки эдамского сыра.

Следствие 5 [22]. Семейство из m гиперплоскостей общего положения рассекает n-мерное пространство на $\sum_{i=0}^{n} C_m^i$ областей.

Разумеется, нижняя оценка теоремы 8 справедлива и для центрированного расположения гиперплоскостей. Поэтому, опираясь на неё, оказывается возможным продвинуться в вопросе о числе пороговых функций.

Число пороговых функций. Покажем теперь, что верхняя оценка, содержащаяся в (5) и (11–12), является точной асимптотикой логарифма числа пороговых функций. Конструктивно построить $2^{n^2(1-o(1))}$ пороговых функций не удалось, несмотря на интенсивные исследования в этой области на протяжении нескольких десятилетий. Этот результат был получен неконструктивно методами, опирающимся на свойства случайных (± 1)-матриц.

В дальнейшем матрицы $\mathbf{A}(m,n)$, компоненты которых независимо и равновероятно принимают значения 1 или -1, будем называть *случайными* (± 1)-*матрицами*. Каждая матрица при этом возникает с вероятностью 2^{-nm}.

Американский учёный Андрей Одлыжко (A. M. Odlyzko), занимаясь задачей, возникшей в теории нейронных сетей, в 1988 году получил результат, оказавшийся решающим для оценки числа пороговых функций. Он установил, что в случайной (± 1)-матрице $\mathbf{A}(m,n)$ число строк m может лишь незначительно уступать числу её столбцов, так что $m \sim n$, $n \to \infty$, но при этом в линейной оболочке её строк с вероятностью, стремящейся к единице, не будет ни одного (± 1)-вектора, отличного от строк матрицы $\mathbf{A}(m,n)$ и им противоположных. Чтобы оценить этот результат, стоит заметить, что строки квадратной (± 1)-матрицы $\mathbf{A}(n,n)$ при $n \to \infty$ с вероятностью, стремящейся к единице, образуют базис, и поэтому в их линейной оболочке содержатся вообще все векторы и, в частности, все (± 1)-векторы. Приведём здесь лишь формулировку этого результата в виде леммы, которая в качестве теоремы 10 будет доказана в Дополнении 2.

Лемма. Существует такая бесконечно малая $\alpha(n) = o(1)$, $n \to \infty$, что вероятность того, что в линейной оболочке строк случайной (± 1)-матрицы $\mathbf{A}(m,n)$, где $m = n(1-\alpha(n))$, содержится ещё хотя бы один (± 1)-вектор, отличный от строк матрицы $\mathbf{A}(m,n)$ и им противоположных, стремится к нулю с ростом n.

Опираясь на этот результат, оказывается возможным получить асимптотику логарифма числа пороговых функций.

Теорема 9 (Ю. А. Зуев, 1989). Для логарифма числа пороговых функций справедливо асимптотическое равенство $\log_2 N_n \sim n^2$, $n \to \infty$.

Доказательство. Предыдущими рассуждениями установлено, что число пороговых функций равно числу многогранных открытых конусов, на которые евклидово $(n+1)$-мерное пространство $< a_0, a_1, \ldots, a_n >$ разбивается гиперплоскостями (10).

Рассмотрим конечное вероятностное пространство, равновероятными точками которого являются (± 1)-матрицы $\mathbf{A}(m, n+1)$, где число строк $m = m(n) = (n+1)\big(1-\alpha(n+1)\big)$ выбрано в соответствии с условием леммы. Обозначим множество этих матриц через $\mathfrak{A}_{n+1}$, $|\mathfrak{A}_{n+1}| = 2^{m(n+1)}$. Будем считать строки матрицы $\mathbf{A}(m,n+1) \in \mathfrak{A}_{n+1}$ нормальными векторами гиперплоскостей (10). Тогда подпространство-пересечение m гиперплоскостей, задаваемых строками матрицы $\mathbf{A}(m,n+1)$ как нормалями, будет ортогональным дополнением линейной оболочки её строк. Поэтому в соответствии с теоремой 8 число конусов можно оценить снизу, подсчитав число различных линейных оболочек, порождаемых матрицами из $\mathfrak{A}_{n+1}$.

Обозначим через $\mathfrak{A}'_{n+1} \subseteq \mathfrak{A}_{n+1}$ подмножество матриц, не содержащих пар одинаковых или противоположных строк. Так как вероятность события, что матрица имеет две одинаковые или две противоположные строки, не превышает $C_m^2 \cdot 2^{-n} \to 0$, имеем $|\mathfrak{A}'_{n+1}| \sim |\mathfrak{A}_{n+1}| \sim 2^{m(n+1)}$, $n \to \infty$.

Обозначим через $\mathfrak{A}''_{n+1} \subseteq \mathfrak{A}_{n+1}$ подмножество тех матриц $\mathbf{A}(m, n+1)$, линейные оболочки которых не содержат (± 1)-векторов, отличный от строк матрицы $\mathbf{A}(m, n+1)$ и им противоположных. Согласно лемме имеем $|\mathfrak{A}''_{n+1}| \sim |\mathfrak{A}_{n+1}| \sim 2^{m(n+1)}$, $n \to \infty$.

Рассмотрим теперь множество $\mathfrak{A}'''_{n+1} = \mathfrak{A}'_{n+1} \cap \mathfrak{A}''_{n+1}$. По-прежнему, имеем $|\mathfrak{A}'''_{n+1}| \sim 2^{m(n+1)}$. Разобьём $\mathfrak{A}'''_{n+1}$ на классы эквивалентности следующим образом. В один класс включим матрицы, которые могут быть получены друг из друга перестановкой строк и заменой некоторых из строк на им противоположные. При этом в каждом классе будет ровно $m!2^m$ матриц, так как $\mathfrak{A}'''_{n+1} \subseteq \mathfrak{A}'_{n+1}$.

Ясно, что все матрицы из одного класса порождают одну и ту же линейную оболочку. Если же взять пару матриц из различных классов, то в каждой из них будет строка, которой и противоположной которой нет в другой матрице. Из условия $\mathfrak{A}'''_{n+1} \subseteq \mathfrak{A}''_{n+1}$, поэтому, вытекает, что линейные оболочки строк этих матриц различны. Таким образом, число различных линейных оболочек, порождаемых матрицами из $\mathfrak{A}'''_{n+1}$, совпадает с числом классов. Отсюда получаем, что число линейных оболочек, а, следовательно, конусов и пороговых функций, асимптотически не меньше чем

$$N_n \geq \frac{2^{m(n+1)}}{m!2^m}.$$

Этим и устанавливается асимптотика логарифма числа пороговых функций

$$\log_2 N_n \geq \log_2 \frac{2^{m(n+1)}}{m!2^m} \sim n^2$$

или

$$N_n = 2^{n^2(1+o(1))}.$$ □

Вопросы для самопроверки

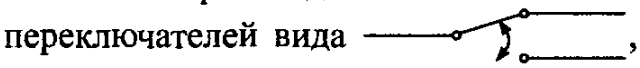
1. Пороговая функция $f(x_1, x_2, x_3, x_4) = x_1x_2x_3 \vee x_1x_2x_4 \vee x_1x_3x_4$ может быть задана неравенством

   а) $x_1 + x_2 + x_3 + x_4 \leq 2$; б) $x_2 + x_3 + x_4 \leq 1$; в) $2x_1 + x_2 + x_3 + x_4 \leq 3$.

2. Вектор параметров Чоу мажоритарной функции от 3 переменных равен
   а) $(1,1,1)$; б) $(2,2,2)$; в) $(3,3,3)$.
3. Пороговое число функции $x_1 \oplus x_2 \oplus x_3$ равно
   а) 2; б) 3; в) 4.

Ответы: 1 — в, 2 — в, 3 — в.

Задачи для самостоятельного решения

1. Нарисовать схему электрической цепи, в которой включение и выключение лампочки производилось бы независимо с помощью каждого из двух переключателей вида  , т. е. изменение положения каждого из переключателей зажигало бы свет, если его до этого не было и гасило бы его, если он горел. Если, закодировав положения первого переключателя символами 0 и 1, связать с ним булеву переменную x_1, а со вторым переключателем — x_2, то какую булеву функцию $f(x_1,x_2)$ реализует схема, если считать, что $f=1$, если свет горит, и $f=0$, если не горит.
2. Совет в составе 4 членов, один из которых является председателем, принимает или отвергает проекты, причём каждый член совета должен высказаться «за» или «против». Решение принимается или отвергается большинством голосов, а в случае равенства решающим является голос председателя. Написать д. н. ф. для соответствующей булевой функции.
3. «Вернувшись домой, Мегрэ позвонил на набережную Орфевр.

   — Говорит Мегрэ. Есть новости?

   — Да, шеф. Поступили сообщения от инспекторов. Жуссье считает, что либо Этьен убийца, либо Франсуа не был пьян и убийство произошло после полуночи. Инспектор Люка просил передать вам, что если убийство произошло после полуночи, то либо Этьен убийца, либо Франсуа лжёт.

   — Немедленно задержите Этьена.»

   Проверьте логичность действий комиссара Мегрэ, принимая во внимание наряду с полученной от инспекторов информацией также известный Мегрэ факт, что трезвый Франсуа никогда не лжёт.
4. Трое подозреваемых Браун, Джонс и Смит дают следующие показания:

   Браун: Джонс виновен, а Смит невиновен;

   Джонс: Если Браун виновен, то виновен и Смит;

   Смит: Я невиновен, но хотя бы один из них (Браун или Джонс) виновен.

   а) Совместимы ли показания троих подозреваемых, т. е. могут ли они быть все верны?

   б) Показания какого из обвиняемых следуют из показаний другого?

в) Если все трое невиновны, то кто совершил лжесвидетельство?

г) Если все показания верны, то кто виновен, а кто невиновен?

д) Если невиновный говорит правду, а виновный лжёт, то кто виновен, а кто невиновен?

5. Чему равно число симметрических (не изменяющихся при любой подстановке аргументов) булевых функций от n переменных?
6. Представить мажоритарную функцию от 3 переменных с помощью полинома Жегалкина.
7. Булева функция $f(x_1, x_2, x_3)$ задана следующим образом: $f(0,0,0) = f(1,1,1) = 0$, а на остальных шести наборах она равна единице. Сколько тупиковых д. н. ф. имеет данная функция? Сколько среди них минимальных?
8. Обобщить решение задачи 7 на случай n переменных: $f(0,\ldots,0) = f(1,\ldots,1) = 0$, а на остальных $2^n - 2$ наборах функция равна единице. Найти все её минимальные д. н. ф. (С. В. Яблонский).
9. Пусть функция $f(x_1, x_2, x_3)$ задана как в задаче 7. Найти число тупиковых д. н. ф. функции $f(x_1, x_2, x_3) \& (x_4 \oplus x_5 \oplus \ldots \oplus x_n)$ (Ю. И. Журавлёв, 1962).
10. Показать что произвольная д. н. ф. булевой функции может быть преобразована в её сокращённую д. н. ф. с помощью алгоритма, применяющего в произвольном порядке следующие две операции над элементарными конъюнкциями:

    а) удаление конъюнкции, поглощаемой другой конъюнкцией, $K_1 \vee K_1 K_2 = K_1$;

    б) обобщённое склеивание $xK_1 \vee \bar{x}K_2 = xK_1 \vee \bar{x}K_2 \vee K_1 K_2$, осуществляемое в случае, если конъюнкция $K_1 K_2$ не поглощается никакой другой входящей в д. н. ф. конъюнкцией.

    Алгоритм заканчивает работу, когда не может быть применена ни одна из двух операций (A. Blake, 1937).
11. Используя формулу включения и исключения, найти число булевых функций от n переменных, существенно зависящих от всех своих переменных. Показать, что почти все булевы функции существенно зависят от всех своих переменных.
12. Доказать неравенство $|\tilde{\alpha} \oplus \tilde{\beta}| \geq |\tilde{\alpha}| - |\tilde{\beta}|$ и найти для него условия равенства (здесь $\tilde{\alpha} \oplus \tilde{\beta}$ обозначает набор, полученный поэлементным сложением по модулю 2 наборов $\tilde{\alpha}$ и $\tilde{\beta}$).
13. Пусть $\tilde{\alpha}$ и $\tilde{\beta}$ — двоичные наборы длины n с расстоянием Хэмминга между ними $d(\tilde{\alpha}, \tilde{\beta}) = r$. Сколько существует наборов $\tilde{\gamma}$ таких, что $d(\tilde{\alpha}, \tilde{\gamma}) + d(\tilde{\gamma}, \tilde{\beta}) = r$?

14. Пусть $\tilde{\alpha}$ и $\tilde{\beta}$ — двоичные наборы длины n с расстоянием Хэмминга между ними $d(\tilde{\alpha},\tilde{\beta})=r$. Сколько существует наборов $\tilde{\gamma}$ таких, что $d(\tilde{\alpha},\tilde{\gamma})=s$ и $d(\tilde{\beta},\tilde{\gamma})=t$?
15. Пусть $\tilde{\alpha},\tilde{\beta}\in B_i^n$ (i-й слой куба), $d(\tilde{\alpha},\tilde{\beta})=2$. Сколько существует наборов $\tilde{\gamma}\in B_i^n$ таких, что $d(\tilde{\gamma},\tilde{\alpha})=2$ и $d(\tilde{\gamma},\tilde{\beta})=2$.
16. Пусть $\tilde{\alpha}\in B_i^n$. Чему равна максимальная мощность множества наборов $\tilde{\beta}\in B_i^n$ таких, что $d(\tilde{\alpha},\tilde{\beta})=2$ и для любых двух различных наборов $\tilde{\beta}_1$, $\tilde{\beta}_2$ этого множества $d(\tilde{\beta}_1,\tilde{\beta}_2)=4$?
17. Пусть $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{\gamma}$ — двоичные наборы, причём $d(\tilde{\alpha},\tilde{\beta})=d(\tilde{\alpha},\tilde{\gamma})=$ $=d(\tilde{\beta},\tilde{\gamma})=r$. Доказать, что r — число чётное и существует в точности один набор $\tilde{\delta}$ такой, что $d(\tilde{\alpha},\tilde{\delta})=d(\tilde{\beta},\tilde{\delta})=d(\tilde{\gamma},\tilde{\delta})=r/2$.
18. Проверить, является ли указанная система функций полной и образует ли она базис в P_2.
    а) $\{x\to y,\ x\oplus y,\ x\vee y\}$;
    б) $\{x\oplus y\oplus z,\ x\vee y,\ 0,\ 1\}$;
    в) $\{x\oplus y\oplus yz,\ x\oplus y\oplus 1\}$;
    г) $\{xy\vee z,\ xy\oplus z,\ xy\sim z\}$.
19. Перечислить все функции $f(x_1,\ldots,x_n)$, принадлежащие одновременно всем классам Поста.
20. Найти число функций $f(x_1,\ldots,x_n)$, входящих хотя бы в один из классов Поста.
21. Доказать, что функция f является шефферовой тогда и только тогда, когда $f\notin \mathrm{T}_0\cup\mathrm{T}_1\cup\mathrm{S}$.
22. Найти число шефферовых функций от n переменных.
23. Показать, что штрих Шеффера и стрелка Пирса являются единственными шефферовыми функциями от двух переменных.
24. Тождественно истинностная логическая формула называется *тавтологией*. Пусть задана некоторая д. н. ф. $D(x_1,\ldots,x_n)$ от n переменных. Вынося x_n за скобку из всех тех элементарных конъюнкций, куда входит x_n, и вынося $\bar{x}_n$ из тех, куда входит $\bar{x}_n$, представим данную д. н. ф. в виде $D(x_1,\ldots,x_n)=x_nA\vee\bar{x}_nB\vee C$, где A, B, C — д. н. ф. от переменных x_1, ..., x_{n-1}. Показать, что $D(x_1,\ldots,x_n)$ является тавтологией в том и только в том случае, если формула $AB\vee C$ является тавтологией. Это позволяет сводить задачу о том, является ли данная д. н. ф. тавтоло-

гией, к аналогичной задаче с меньшим числом переменных (M. Davis, H. Putnam, 1960).

К. н. ф. называется *выполнимой*, если существует набор значений переменных, на котором она принимает значение «истина». Пользуясь соображениями двойственности, сформулировать аналогичную уменьшающую число переменных процедуру для проверки свойства выполнимости.

25. Дать однородное пороговое представление $f:\{-1,1\}^4 \to \{-1,1\}$ для булевой функции $f(y_1, y_2, y_3, y_4)$ в задаче 2.
26. Доказать, что для равенства порогового числа монотонной булевой функции длине её кратчайшей д. н. ф. необходимо и достаточно выполнения для каждой пары $\tilde{\alpha}_i$, $\tilde{\alpha}_j$ её нижних единиц следующих двух неравенств $|\tilde{\alpha}_i| > |\tilde{\alpha}_i \& \tilde{\alpha}_j| + 1$ и $|\tilde{\alpha}_j| > |\tilde{\alpha}_i \& \tilde{\alpha}_j| + 1$ (здесь $\tilde{\alpha}_i \& \tilde{\alpha}_j$ обозначает набор, полученный поэлементной конъюнкцией наборов $\tilde{\alpha}_i$ и $\tilde{\alpha}_j$, а $|\tilde{\alpha}_i|$ — число единиц в наборе $\tilde{\alpha}_i$).
27. Пусть $M = \{(-1, 1, 1, \ldots, 1), (-1, -1, 1, \ldots, 1), \ldots (-1, -1, -1, \ldots, -1)\}$ — множество из n ± 1-векторов в евклидовом n-мерном пространстве, n — чётное число. Доказать, что для любого n-мерного ± 1-вектора в M найдётся ортогональный ему вектор.
28. Множество точек находится в общем положении в евклидовом n-мерном пространстве, если никакие $i+2$ точки множества не лежат в i-мерном аффинном подпространстве. Сколько линейно отделимых подмножеств имеет в n-мерном пространстве множество из K точек общего положения?

Литература

1. *Васильев Ю. Л., Глаголев В. В.* Метрические свойства дизъюнктивных нормальных форм // Дискретная математика и математические вопросы кибернетики / Под ред. С. В. Яблонского и О. Б. Лупанова. М.: Наука, 1974. С. 99–148.
2. *Гаврилов Г. П., Сапоженко А. А.* Задачи и упражнения по курсу дискретной математики. М.: Наука, 2004.
3. *Глухов М. М., Козлитин О. А., Шапошников В. А., Шишков А. Б.* Задачи и упражнения по математической логике, дискретным функциям и теории алгоритмов. М.: Лань, 2008.
4. *Журавлёв Ю. И.* Алгоритмы построения минимальных дизъюнктивных нормальных форм для функций алгебры логики // Дискретная математика и математические вопросы кибернетики / Под ред. С. В. Яблонского и О. Б. Лупанова. М.: Наука, 1974. С. 67–98.
5. *Журавлёв Ю. И., Флёров Ю. А., Федько О. С., Дадашев Т. М.* Сборник задач по дискретному анализу. М.: МФТИ, 2004.
6. *Зуев Ю. А.* Комбинаторно-вероятностные и геометрические методы в пороговой логике // Дискретная математика. 1991. Т. 3. Вып. 2. С. 47–57.

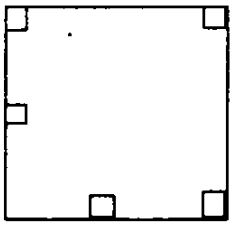
7. *Зуев Ю. А.* Пороговые функции и пороговые представления булевых функций // Математические вопросы кибернетики. Вып. 5. М.: Наука, 1994. С. 5–61.
8. *Зуев Ю. А., Иванов С. К.* Обучение и самообучение в процедурах взвешенного голосования // Журнал вычислительной математики и математической физики. 1995. Т. 35. № 1. С. 104–121.
9. *Клейтмен Д.* О проблеме Дедекинда: число монотонных булевых функций // Кибернетический сборник. Новая серия. Вып. 7. М: Мир, 1970. С. 43–52.
10. *Клини С. К.* Математическая логика. 4-е изд. М.: Издательство ЛКИ/URSS, 2008.
11. *Коршунов А. Д.* О числе монотонных булевых функций // Проблемы кибернетики. Вып. 38. М.: Наука, 1981. С. 5–108.
12. *Липкин Л. И.* О представлении булевых функций с заданным числом нулей системами линейных неравенств // Журнал вычислительной математики и математической физики. 1987. Т. 27. № 6. С. 949–951.
13. *Мендельсон Э.* Введение в математическую логику. М.: Наука, 1991.
14. *Нигматуллин Р. Г.* Некоторые метрические соотношения в единичном кубе // Дискретный анализ. Вып. 9. Новосибирск, 1967. С. 47–58.
15. *Нигматуллин Р. Г.* Сложность булевых функций. М.: Наука, 1991.
16. *Сапоженко А. А.* Проблема Дедекинда и метод граничных функционалов. М.: Наука, 2009.
17. *Столл Р. Р.* Множества. Логика. Аксиоматические теории. М.: Просвещение, 1968.
18. *Чухров И. П.* О тупиковых комплексах граней в единичном кубе // Дискретная математика. 2011. Т. 23. Вып. 1. С. 132–158.
19. *Шабанин О. В.* О сложности дизъюнктивной нормальной формы пороговых функций // Дискретная математика. 2000. Т. 12. Вып. 2. С. 85–92.
20. *Яблонский С. В., Гаврилов Г. П., Кудрявцев В. Б.* Функции алгебры логики и классы Поста. М.: Наука, 1966.
21. *Яблонский С. В.* Введение в дискретную математику. М.: Наука, 1986.
22. *Buck R. C.* Partition of Space // Amer. Math. Monthly. 1943. Vol. 50. № 9. С. 541–544.
23. *Muroga S.* Threshold logic and its application. N. Y.: Wiley, 1971.
24. *Winder R. O.* Partitions of N-space by hyperplanes // SIAM J. Appl. Math. 1966. Vol. 14. № 4. С. 811–818.

Комментарии к литературе

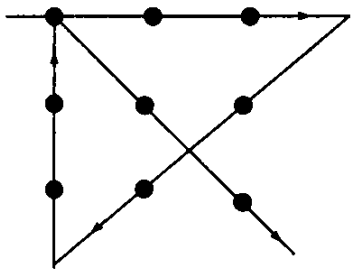
Хорошим элементарным введением в теорию множеств и логику высказываний является [17], [10] — намного более серьёзный учебник, включающий также и философские экскурсы автора, [12] — общепризнанный классический курс математической логики. В [4] прекрасно представлены элементарные методы упрощения д. н. ф., а в [1] приведены важные для понимания трудности решения этой задачи типичные и экстремальные оценки сложности д. н. ф. Монография [15] содержит много дополнительной информации о булевых функциях и д. н. ф. Проблеме оценки числа монотонных булевых функций посвящены [11] и [16]. Больше узнать о результатах Поста можно в [20], а в [21] можно ознакомиться с обобщением булевых функций — функциями k-значной логики. Значительное число задач, относящихся к булевым функциям, заинтересованный читатель найдёт в [2], [3] и [5]. Монография [23] является энциклопедией результатов, полученных в пороговой логике до 1971 года. С более поздними результатами в этой области можно ознакомиться в [7]. В [8] рассматриваются алгоритмы обучения и самообучения порогового элемента.

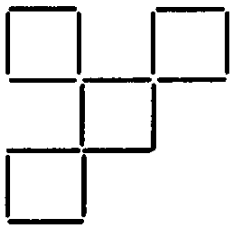
Ответы и указания к решению задач

Глава 0

0.1. 

0.2. 59 секунд.

0.3. 

0.4. 

0.5. 10.

0.6. Нет, так как поля a1 и h8 одного цвета.

0.7. Нужно вынуть шар из урны с этикеткой «белый и чёрный».

0.9. При первом взвешивании на каждую чашу следует положить по три монеты.

0.9. При первом взвешивании на каждую чашу одних из весов кладётся по четыре монеты. В случае равенства по три монеты с каждой чаши кладётся на чаши других весов. В случае равенства на этих же весах взвешиваются две монеты, оставшиеся на чашах первых весов.

0.10. Можно одновременно завесить одну монету из первого мешка, две монеты из второго и т. д., десять монет из десятого.

0.11. Если двое мужчин не присутствовали одновременно на рауте, то все женщины должны были присутствовать там в промежутке между посещениями этих мужчин.

0.12. Утверждение № 2.

0.13. Рост выбранного первым не меньше роста второго.

0.14. Наполняем сосуд 3 л и переливаем из него воду в сосуд 5 л. Затем снова из большого сосуда наполняем сосуд 3 л и наполняем из него сосуд 5 л. Теперь в большом сосуде 2 л, в среднем — 5 л и в малом — 1 л. Выливаем воду из среднего сосуда в большой сосуд и наливаем в средний сосуд 1 л из малого сосуда. Затем из большого сосуда наполняем малый и переливаем из него воду в средний сосуд.

0.15. Одинаково, так как остались неизменными объёмы жидкостей в бокале и в бочке.

0.16. Вы кладёте документ в коробочку и запираете её на свой замочек. Получив коробочку, ваш друг запирает её на свой замочек и посыла-

ет вам. Получив коробочку, вы снимаете свой замочек и посылаете её другу.

0.17. Каждый раз нужно брать столько камней, чтобы число оставшихся камней было кратно четырём.

0.18. Вначале идут первый и второй, затем первый возвращается и передаёт фонарик третьему и четвёртому, которые, перейдя мост, передают фонарик второму, который возвращается, а затем переходит мост вместе с первым.

0.19. Индукция по n.

0.20. Индукция по n:

$$9^{n+2}-8(n+1)-9=9(9^{n+1}-8n-9)+64(n+1).$$

0.21. $a^{10}=\left(\left(a^2\right)^2\right)^2\cdot a^2$.

0.22. Каждый простой сомножитель входит в $(m,n)\cdot[m,n]$ и mn с одинаковой кратностью.

0.23. 30 лошадей и 40 быков или 51 лошадь и 9 быков.

0.24. На отрезке натурального ряда $\{1,2,\ldots,pq\}$ имеется $pq/p=q$ чисел, кратных p, и $pq/q=p$ чисел, кратных q, при этом одно число, а именно pq, кратно обоим этим числам. Поэтому $\varphi(pq)=pq-p-q+1=(p-1)(q-1)$.

0.25. Имеем $a-b=kp=lq$. Теперь из единственности разложения на простые множители следует, что $q\mid k$, поэтому $(pq)\mid(a-b)$.

0.26. Если $n=kl$, где k — нечётно, то $2^n+1=(2^l)^k+1^k$ делится на 2^l+1.

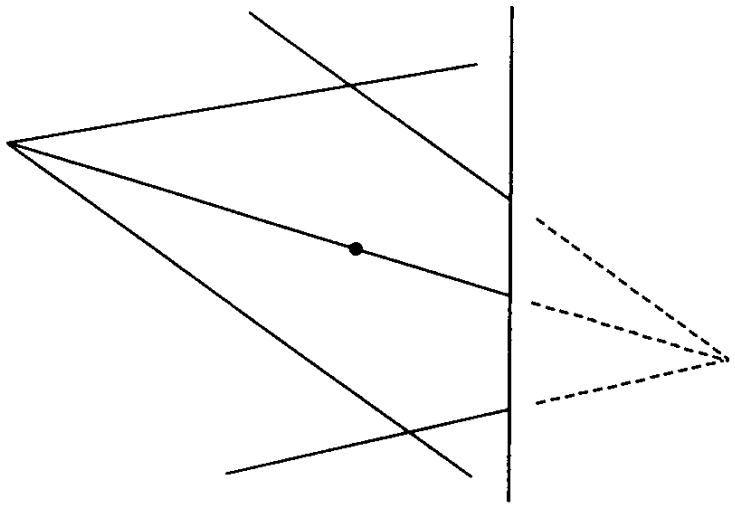
0.27. Если $q\mid 2^{2^k}+1$, то $2^{2^k}\equiv -1\,(\mathrm{mod}\,q)$. Тогда $2^{2^{k+l}}=\left(2^{2^k}\right)^{2^l}\equiv 1(\mathrm{mod}\,q)$, а $2^{2^{k+l}}+1\equiv 2(\mathrm{mod}\,q)$, и q не делит $2^{2^{k+l}}+1$.

0.28. При повороте n-угольника $A_1A_2\ldots A_n$ вокруг его центра O на угол $2\pi/n$ вектор суммы $\overrightarrow{OA_1}+\overrightarrow{OA_2}+\ldots+\overrightarrow{OA_n}$ также повернётся на угол $2\pi/n$, но при этом останется неизменным, так как n-угольник перейдёт в себя. Отсюда следует, что $\overrightarrow{OA_1}+\overrightarrow{OA_2}+\ldots+\overrightarrow{OA_n}=\vec{0}$.

0.29. Пусть $\alpha_1\mathbf{a}_1+\alpha_2\mathbf{a}_2+\ldots+\alpha_k\mathbf{a}_k=\mathbf{0}$. Помножив это равенство скалярно на $\mathbf{a}_1$, получаем $\alpha_1(\mathbf{a}_1,\mathbf{a}_1)=0$, откуда $\alpha_1=0$. Аналогично доказывается равенство нулю и всех других коэффициентов.

0.30. Пусть $\mathbf{v}\in(S+T)^{\perp}$. Тогда $\mathbf{v}\in S^{\perp}$ и $\mathbf{v}\in T^{\perp}$. Следовательно $\mathbf{v}\in S^{\perp}\cap T^{\perp}$. Поэтому $(S+T)^{\perp}\in S^{\perp}\cap T^{\perp}$. Пусть теперь $\mathbf{v}\in S^{\perp}\cap T^{\perp}$. Тогда $\mathbf{v}\in S^{\perp}$ и $\mathbf{v}\in T^{\perp}$. Следовательно, $(\mathbf{v},\mathbf{s}+\mathbf{t})=0$ и $\mathbf{v}\in(S+T)^{\perp}$. Поэтому $S^{\perp}\cap$ $\cap T^{\perp}\in(S+T)^{\perp}$. Следовательно $(S+T)^{\perp}=S^{\perp}\cap T^{\perp}$.

0.31. Можно воспользоваться центральной симметрией относительно заданной точки, при которой образы заданных прямых пересекаются в пределах чертежа, а требуемая прямая переходит в себя.



0.32. Каждая подгруппа циклической группы порождается наименьшей из степеней входящих в неё элементов.

0.33. Каждая транспозиция, переставляющая два соседних элемента, может быть получена как $(i,i+1)=(123\ldots n)^{n-i+1}(12)(123\ldots n)^{i-1}$. Произвольная транспозиция может быть получена с помощью таких транспозиций как

$$(ij)=[(i,i+1)(i+1,i+2)\ldots(j-1,j)][(j-1,j-2)(j-2,j-3)\ldots(i,i+1)].$$

Произвольная подстановка разлагается в произведение независимых циклов, каждый из которых может быть получен с помощью транспозиций как $(i_{j_1}i_{j_2}\ldots i_{j_k})=(i_{j_{k-1}}i_{j_k})(i_{j_{k-2}}i_{j_{k-1}})\ldots(i_{j_1}i_{j_2})$.

0.34. Если a — фиксированный элемент подмножества, а b_i пробегает все элементы подмножества, то ab_i также пробегает все элементы подмножества. Поэтому существует элемент b_i такой, что $ab_i=a$, т. е. единичный. Теперь аналогичным образом может быть доказана и принадлежность подмножеству элемента a^{-1}.

0.35. Для произвольных элементов группы a и b имеем $(ab)(ab)=e$. Помножая слева на a, получаем $bab=a$. Помножая слева на b, получаем $ab=ba$.

0.36. Пусть k — порядок элемента ab. Имеем $(ab)^k = a^k b^k = e$, откуда $a^k = b^{-k} = \left(b^{-1}\right)^k$. Порядок элемента b^{-1} совпадает с порядком элемента b и равен n. Но единственным общим элементом двух циклических подгрупп, порядки которых взаимно просты, является единичный элемент, так как в противном случае отличный от единицы порядок подгруппы их пересечения был бы общим делителем порядков подгрупп. Поэтому $k = mn$.

0.37. Обозначив соответствующие множества через A, B, C и D, а множество всех студентов потока через U, имеем

$$\max|A\cap B\cap C\cap D| = \min\{|A|,|B|,|C|,|D|\} = 70\%;$$

$$\min|A\cap B\cap C\cap D| = \min|\overline{\overline{A\cap B\cap C\cap D}}| = \min|\overline{\overline{A}\cup\overline{B}\cup\overline{C}\cup\overline{D}}| =$$

$$= |U| - \max|\overline{A}\cup\overline{B}\cup\overline{C}\cup\overline{D}| = |U| - \left(|\overline{A}|+|\overline{B}|+|\overline{C}|+|\overline{D}|\right) = 10\%.$$

0.38. $n(n-3)/2$.

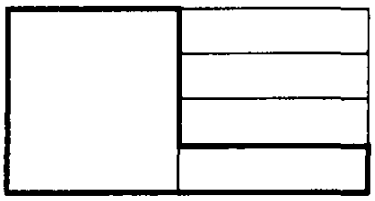
0.39. $9\cdot 10^3$

0.40. $12\cdot 3! = 72$.

0.41. $n!(n-1)!$ способов, так как мужчин можно рассадить вокруг стола $(n-1)!$ способами, и $n!$ способами между ними можно посадить женщин.

0.42. Полное число элементарных исходов равно числу перестановок 9 карточек, т. е. 9!, благоприятных среди них — 4!, так как перестановка букв *а* между собой не изменяет слова *катамаран*. Поэтому искомая вероятность есть $4!/9! = 1/15\,120$.

0.43.



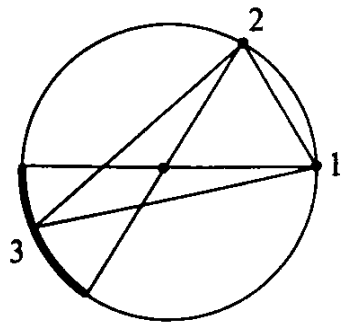
$$P = \frac{1/8}{1/2+1/8} = \frac{1}{5}.$$

0.44. $(1/2)^8 = 1/256$.

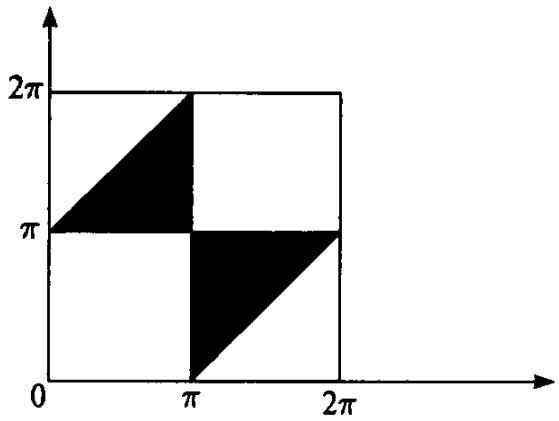
0.45. При продолжении игры возможно не более 2 подбрасываний монетки, имеющих 4 возможных исхода: $(00),(01),(10),(11)$, три из которых благоприятствуют Вове. Поэтому вероятность победы Вовы равна $3/4$, а Пети — $1/4$.

0.46. $P = 1/4$. Как видно из рис. а), для покрытия треугольником центра окружности необходимо и достаточно, чтобы третья точка оказалась внутри угла, вертикального с углом, образованным первыми двумя точ-

ками. Отсюда, принимая первую точку за начало отсчёта и откладывая по осям координат положения двух других точек, получаем выделенную область на рис. б), соответствующую покрытию. Её площадь составляет $1/4$ часть от площади всего квадрата (см. также задачу Д2.9)

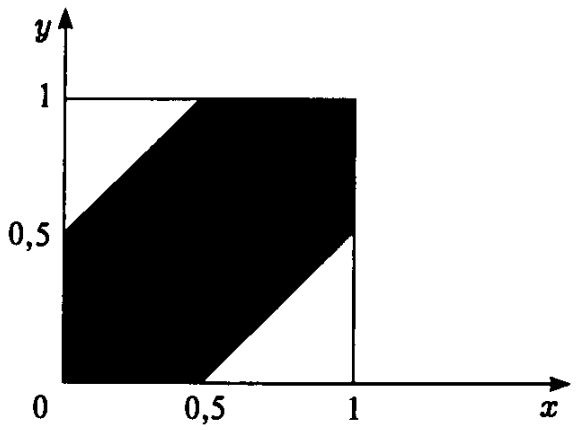


а)



б)

0.47. Приход первого из встречающихся в момент времени $6+x$, а второго — в момент времени $6+y$ задаётся точкой (x,y) внутри единичного квадрата: $0 \le x \le 1$, $0 \le y \le 1$. Условие же встречи задаётся соотношением $|x-y| \le 0{,}5$, которому удовлетворяет выделенная на рисунке область. Её площадь равна $3/4$, что и есть искомая вероятность.



0.48. Если n — число рыб в озере, то вероятность выловить помеченную рыбу есть $p = n_1/n$. По закону больших чисел отношение m/n_2 с ростом n стремится к p. Приравнивая два отношения, получаем $n = n_1 n_2/m$.

0.49. Пусть n — полное число ошибок, p_1, p_2 — вероятности обнаружения ошибки корректорами. Считая эти события независимыми, получаем, что вероятность обнаружения ошибки обоими корректорами есть $p_1 p_2$. Опираясь на закон больших чисел, полагаем $np_1 = n_1$, $np_2 = n_2$, $np_1 p_2 = n_{12}$. Умножая почленно два первых уравнения и деля на третье, получаем $n = n_1 n_2/n_{12}$, а число незамеченных ошибок есть $n_1 n_2/n_{12} - (n_1 + n_2 - n_{12})$.

0.50. Минимальное число отрядов, с помощью которого всегда достижим нужный результат, равно четырём.

0.51. Достаточно приписать рёбрам графа в качестве весов различные простые числа, а затем каждой вершине поставить в соответствие число, равное произведению весов инцидентных ей рёбер.

0.52. Если в компании двое или более лиц не имеют знакомых, то среди них можно выбрать нужную пару. В противном случае, удалив, максимум, одного человека, получим подкомпанию из n человек, в которой число знакомых каждого варьируется от 1 до $n-1$ и, следовательно, найдутся двое, имеющие одинаковое число знакомых.

0.53. У произвольного человека группы среди остальных 5 человек имеются трое знакомых ему или трое незнакомых. В первом случае, если среди троих его знакомых нет знакомых между собой, то имеем попарно незнакомую тройку, а если среди них найдутся двое знакомых — то попарно знакомую тройку. Аналогично рассматривается второй случай.

0.54. Если у какого-то члена группы число его знакомых, находящихся в одной с ним комнате, превышает число его знакомых из другой комнаты, то перемещаем его в другую комнату. При этом общее число знакомых пар, разнесённых по разным комнатам, возрастает. Этим гарантируется, что за конечное число подобных перемещений будет достигнуто требуемое размещение.

0.55. Индукция по n. При проведении $(n+1)$-й плоскости число трёхмерных областей, на которые разбивается пространство, возрастает на число плоских областей, на которые эта плоскость разбивается n прямыми, а это число по доказанному во Вводной главе равно $(n^2+n+2)/2$. Поэтому число областей, на которые пространство разбивается $n+1$ плоскостями, равно $(n^3+5n+6)/6+(n^2+n+2)/2=$ $=\left((n+1)^3+5(n+1)+6\right)/6$ (см также теорему 2.8.6).

Глава 1

1.1. $A_5^4 \cdot A_5^2 \cdot 4! = 57\,600.$

1.2. $C_{36}^6 \cdot C_{30}^6 \cdot C_{24}^6 \cdot C_{18}^6 \cdot 12! = \dfrac{36!}{(6!)^4}.$

1.3. а) $C_5^3 \cdot C_3^1 = 30$; б) $C_4^2 \cdot C_4^2/2 = 18$.

1.4. а) 2^{mn}; б) C_{mn}^k; в) $A_{2^n}^m$; г) $(2^n-1)^m$; д) $A_{2^n-1}^m$.

1.5. а) $\dfrac{n!}{n_1!\ldots n_k!}$; б) $\dfrac{n!}{n_1!\ldots n_k!m_1!\ldots m_2!}$; в) $\dfrac{m!}{m_1!\ldots m_k!}$; г) 1.

1.6. Каждые четыре точки из n задают одну точку пересечения хорд внутри окружности. Поэтому искомое число равно C_n^4.

1.7. а) C_{10}^4; б) $\bar{C}_{10}^4 - 1$.

1.8. C_{m+1}^n, так как единицы ставятся между нулями, считая их начало и конец.

1.9. Если первое место оставить свободным, то выбрать множество из n занимаемых мест можно согласно решению предыдущей задачи C_m^n способами. А если первое место занимается, то второе и $(n+m)$-е места должны остаться свободными, и выбрать оставшиеся $n-1$ мест можно C_{m-1}^{n-1} способами. Полное число способов выбора множества занимаемых мест равно, таким образом, $C_m^n + C_{m-1}^{n-1} = \dfrac{m+n}{m} C_m^n$, а число способов рассаживания по ним n человек есть $n!\dfrac{m+n}{m} C_m^n = (m+n)\dfrac{(m-1)!}{(m-n)!}$.

1.10. $\bar{C}_6^5$.

1.11. $\bar{C}_3^2 \cdot \bar{C}_3^4 \cdot \bar{C}_3^5$.

1.12. $\bar{C}_{n+1}^m$.

1.13. Следует из формулы включения и исключения.

1.14. а) $C_{n+1}^{m+1} = C_n^m + C_n^{m+1} = C_n^m + C_{n-1}^m + C_{n-1}^{m+1} =$

$$C_n^m + C_{n-1}^m + C_{n-2}^m + C_{n-2}^{m+1} = \ldots = \sum_{i=m}^{n} C_i^m;$$

б) $\sum\limits_{k=0}^{n}\sum\limits_{r=0}^{k} C_n^k C_k^r = \sum\limits_{k=0}^{n} C_n^k \sum\limits_{r=0}^{k} C_k^r = \sum\limits_{k=0}^{n} C_n^k 2^k = (1+2)^n = 3^n$.

1.15. При $n=1$ неравенства справедливы. Докажем их индукцией по n.

$$C_{2(n+1)}^{n+1} = 2\frac{2n+1}{n+1} C_{2n}^n \geq 2\frac{2n+1}{n+1}\frac{4^n}{2\sqrt{n}} > \frac{2n+1}{n\sqrt{n+1}} 4^n > \frac{1}{2\sqrt{n+1}} 4^{n+1};$$

$$C_{2(n+1)}^{n+1} = 2\frac{2n+1}{n+1} C_{2n}^n \leq 2\frac{2n+1}{n+1}\frac{4^n}{\sqrt{3n+1}} = \frac{2n+1}{2n+2}\frac{4^{n+1}}{\sqrt{3n+1}} < \frac{4^{n+1}}{\sqrt{3n+4}}.$$

1.16. Есть! С помощью формулы включения и исключения число умных, красивых и богатых находится как $20-(11+10+9)+(3+4+4)=1$.

1.17. По формуле включения и исключения, с учётом того, что среди первых n членов натурального ряда число чисел, делящихся одновременно на m и k, равно $[n/l]$, где $l=[m,k]$, имеем

$$1000-[1000/6]-[1000/10]-[1000/15]+[1000/30]+$$
$$+[1000/30]+[1000/30]-[1000/30]=734.$$

1.18. Имеется $(n!)^2$ элементарных исходов. Для нахождения числа благоприятствующих исходов воспользуемся (1.3.8). В случае а) $S_j = C_n^j\left((n-j)!\right)^2$, и искомая вероятность оказывается равной

$$\frac{\sum_{j=l}^{n}(-1)^{j-l}C_j^l C_n^j\left((n-j)!\right)^2}{(n!)^2}=\frac{1}{n!l!}\sum_{j=l}^{n}(-1)^{j-l}\frac{(n-j)!}{(j-l)!}.$$

В случае б) $S_j = C_n^j\left(2(n-j)!n!-\left((n-j)!\right)^2\right)$, и вероятность равна

$$\frac{1}{n!l!}\sum_{j=l}^{n}(-1)^{j-l}\frac{2n!-(n-j)!}{(j-l)!}.$$

1.19. Обозначим через U множество всех рассаживаний, в которых мужчины и женщины чередуются, а через U_i, $i=1,2,\ldots,n$ — множество тех из них, в которых i-я пара супругов сидит рядом, и воспользуемся (1.3.6). Имеем $S_0=|U|=n!(n-1)!$, $|U_i|=2\left((n-1)!\right)^2$, $i=1,2,\ldots,n$ и $S_1=2n!(n-1)!$. Для нахождения общей формулы для S_j выберем j пар (C_n^j способов) и рассадим их вокруг стола как цельные объекты. Таких рассаживаний по кругу существует $(j-1)!$, так как цикл из j элементов можно образовать $(j-1)!$ способами. Зададим количество свободных мест в промежутках между парами для остальных участников приёма ($\overline{C}_j^{2n-2j}=C_{2n-j-1}^{2n-2j}$ способов). Задав теперь порядок супругов в любой из выбранных пар одним из двух возможных способов, получаем однозначно определённый порядок чередования мужчин и женщин за столом, и рассадить остальных участников приёма по намеченным свободным местам можно $\left((n-j)!\right)^2$ способами. Окончательно получаем $S_j=2C_n^j(j-1)!\,C_{2n-j-1}^{2n-2j}\left((n-j)!\right)^2=\frac{2n!}{2n-j}C_{2n-j}^{j}(n-j)!$, откуда число искомых рассаживаний равно $2n!\sum_{j=0}^{n}(-1)^j\frac{(n-j)!}{2n-j}C_{2n-j}^{j}$.

1.20. В производящей функции для числа способов раздачи монет

$$(x+x^2+x^3)^5 = x^5(1+x+x^2)^5 =$$

$$= x^5\left(\frac{1-x^3}{1-x}\right)^5 = x^5(1-x^3)^5\frac{1}{(1-x)^5} = x^5\sum_{i=0}^{5}(-1)^i C_5^i x^{3i}\cdot\sum_{i=0}^{\infty}C_{4+i}^4 x^i$$

коэффициент при x^{12} равен $C_5^0C_{4+7}^4 - C_5^1C_{4+4}^4x^4 + C_5^2C_{4+1}^4 = 30$.

1.21. Подстановка на множестве номеров билетов

$$\begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 & i_6 \\ i_1 & i_2 & i_3 & 9-i_4 & 9-i_5 & 9-i_6 \end{pmatrix}$$

переводит номера с равной суммой первых и последних трёх цифр в номера с суммой всех цифр, равной 27. Остаётся найти число целочисленных неотрицательных решений уравнения $x_1+x_2+x_3+x_4+$ $+x_5+x_6 = 27$. В производящей функции

$$(1+x+x^2+x^3+x^4+x^5+x^6+x^7+x^8+x^9)^6 =$$

$$= \left(\frac{1-x^{10}}{1-x}\right)^6 = \sum_{i=0}^{6}(-1)^i C_6^i x^{10i}\cdot\sum_{i=0}^{\infty}C_{5+i}^5 x^i$$

коэффициент при x^{27} равен $C_6^0C_{5+27}^5 - C_6^1C_{5+17}^5 + C_6^2C_{5+7}^5 = 55\,252$.

1.22. $7C_7^2/C_{28}^2 = 7/18$.

1.23. Среди 11 благоприятствующих событию комбинаций

$$(6,6,6),(6,6,9),(D,9,9),(2,8,T),(B,9,10),(D,7,T),$$

$$(D,8,10),(K,6,T),(K,7,10),(K,8,9),(6,7,8)$$

одна состоит из карт одинакового достоинства и в двух имеются по две карты одного достоинства. В остальных 8 комбинациях все карты различных достоинств. Поэтому искомая вероятность равна

$$\mathrm{P}(A) = \frac{C_4^3 + 2\cdot4\cdot C_4^2 + 8\cdot4^3}{C_{36}^3} = \frac{564}{7140} \approx 0{,}079.$$

1.24. а) $4C_{13}^5/C_{52}^5$; б) $13C_{48}^1/C_{52}^5$; в) $A_{13}^2\cdot C_4^3\cdot C_4^2/C_{52}^5$;

г) $(9\cdot4^5-4\cdot9)/C_{52}^5$; д) $13\cdot C_4^3\cdot C_{12}^2\cdot4^2/C_{52}^5$;

е) $C_{13}^2\cdot(C_4^2)^2\cdot11\cdot4/C_{52}^5$; ж) $13\cdot C_4^2\cdot C_{12}^3\cdot4^3/C_{52}^5$.

1.25. Обозначим число способов через a_n. Первым шагом можно преодолеть одну или две ступеньки. Это даёт рекуррентное соотношение

$a_n = a_{n-1} + a_{n-2}$, которое задаёт последовательность Фибоначчи с начальными членами $a_0 = 1$, $a_1 = 1$, $a_2 = 2$, $a_3 = 3$, $a_4 = 5$, ... Отсюда

$$a_n = \frac{1}{\sqrt{5}}\left(\left(\frac{\sqrt{5}+1}{2}\right)^{n+1} + (-1)^n\left(\frac{\sqrt{5}-1}{2}\right)^{n+1}\right)$$

1.26. $1+n+n(n-1)/2$.

1.27. $2^n - 1$.

1.28. $\begin{cases} a_n = 2^{n+1}(2n+7); \\ b_n = -2^{n+1}(2n+3). \end{cases}$

1.29. $n^2(n+1)^2/4$.

1.30. Подсчитывая в методе траекторий число путей с помощью принципа отражения, получаем, что искомая вероятность есть

$$\left(C_{m+n}^m - C_{m+n}^{m+1}\right)\Big/ C_{m+n}^m = 1 - n/(m+1).$$

1.31. Число способов построения есть число Каталана $C_{2n}^n/(n+1)$, так как эта задача изоморфна задаче о числе ± 1-векторов длины $2n$. Для установления изоморфизма достаточно выстроить всю группу по росту и пометить символом «$+1$» тех, кто встанет в первую шеренгу, и символом «-1» тех, кто встанет во вторую.

1.32. $\dfrac{C_{n+k}^m}{C_n^m} = \dfrac{(n+1)(n+2)\ldots(n+k)}{(n-m+1)(n-m+1)\ldots(n-m+k)} = \prod_{i=1}^{k}\left(1+\dfrac{m}{n-m+i}\right) \sim e^{mk/n}$.

Глава 2

2.1.

2.2. $x_1x_2 \vee x_1x_3 \vee x_1x_4 \vee x_2x_3x_4$.

2.3. Обозначим высказывания: P — «Этьен — убийца», Q — «Франсуа пьян», R — «убийство после полуночи», S — «Франсуа лжёт». Из посылок $P \oplus \bar{Q}R$, $R \to (P \oplus S)$ и $\bar{Q} \to \bar{S}$ вытекает P.

2.4. а) совместимы; б) из показаний Брауна показания Смита; в) Браун и Смит; г) Джонс виновен, а Браун и Смит невиновны; д) Браун и Смит виновны, а Джонс невиновен.

2.5. 2^{n+1}, так на каждом из $n+1$ слоёв симметрическая функция принимает постоянное значение — 0 или 1.

2.6. $x_1x_2 \oplus x_1x_3 \oplus x_2x_3$.

2.7. 5 тупиковых, из них 2 минимальные:

$$x_1\bar{x}_2 \vee x_2\bar{x}_3 \vee x_3\bar{x}_1 \text{ и } x_1\bar{x}_3 \vee x_3\bar{x}_2 \vee x_2\bar{x}_1.$$

2.8. $(n-1)!$ минимальных д. н. ф. вида $x_1\bar{x}_{\pi(1)} \vee x_{\pi(1)}\bar{x}_{\pi^2(1)} \vee \ldots x_{\pi^{n-1}(1)}\bar{x}_1$, где π — произвольная циклическая подстановка на $\{1, 2, \ldots, n\}$.

2.9. $f(x_1, x_2, x_3)$ имеет 5 тупиковых д. н. ф. $D_1, \ldots, D_5$, а единственная д. н. ф. функции $x_4 \oplus x_5 \oplus \ldots \oplus x_n$ состоит из 2^{n-4} конъюнкций K_i, $i = 1, 2, \ldots, 2^{n-4}$ ранга $n-3$, интервалы которых — изолированные вершины куба B^{n-3}. Конъюнктивно присоединяя к каждой из конъюнкций K_i любую из 5 д. н. ф. $D_{j(i)}$, получаем тупиковую д. н. ф. $\bigvee_{i=1}^{2^{n-4}} K_i D_{j(i)}$ функции $f(x_1, x_2, x_3) \& (x_4 \oplus x_5 \oplus \ldots \oplus x_n)$, у которой имеется, таким образом, $5^{2^{n-4}}$ тупиковых д. н. ф.

2.10. Поглощаемые конъюнкции, удаляемые операцией 1), не могут возникнуть вновь в результате операции 2), так как остаются поглощаемыми. Покажем, что, если некоторый простой импликант отсутствует в д. н. ф., то применима операция 2). Тогда из конечности множества элементарных конъюнкций будет следовать, что на некотором шаге алгоритма в результате применений операции 2) будут получены все простые импликанты, а все конъюнкции, не являющиеся простыми импликантами, будут удалены в результате применений операции 1), т. е. будет получена сокращённая д. н. ф.

Пусть простой импликант K отсутствует в д. н. ф. Множество элементарных конъюнкций, поглощаемых K, но не поглощаемых ни одной присутствующей в д. н. ф. конъюнкцией, не пусто, так как в него входит K. Пусть K' — конъюнкция максимального ранга в этом множестве, x — отсутствующая в ней переменная. Тогда конъюнкция $K'x$ поглощается некоторой входящей в д. н. ф. и содержащей x конъюнкцией K_1, а конъюнкция $K'\bar{x}$ — содержащей $\bar{x}$ конъюнкцией K_2, и к этим конъюнкциям применима операция 2).

2.11. Пусть U_i — множество функций, не зависящих от переменной x_i. Тогда в формуле включения и исключения $S_j = C_n^j 2^{2^{n-j}}$ и согласно (1.3.6) число булевых функций, существенно зависящих от n переменных равно $\sum_{j=0}^{n}(-1)^j C_n^j 2^{2^{n-j}}$. Используя неравенство Бон-

феррони, получаем, что таких функций не менее $2^{2^n} - n2^{2^{n-1}} \sim 2^{2^n}$, $n \to \infty$.

2.12. Имеем $|\tilde{\alpha} \oplus \tilde{\beta}| = d(\tilde{\alpha}, \tilde{\beta})$, $|\tilde{\alpha}| = d(\tilde{0}, \tilde{\alpha})$, $|\tilde{\beta}| = d(\tilde{0}, \tilde{\beta})$, и доказываемое неравенство следует из неравенства $d(\tilde{\alpha}, \tilde{\beta}) \geq d(\tilde{0}, \tilde{\alpha}) - d(\tilde{0}, \tilde{\beta})$, вытекающего из неравенства треугольника. Доказываемое неравенство обращается в равенство в том и только в том случае, если $\tilde{\alpha} \geq \tilde{\beta}$.

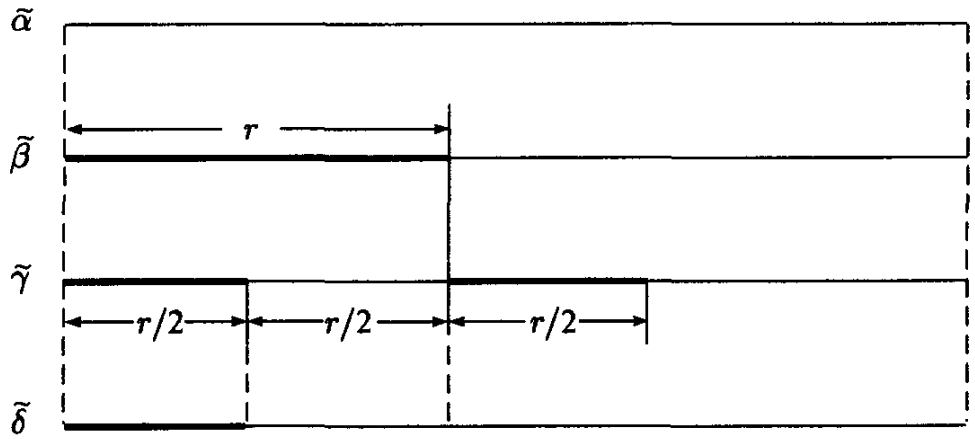
2.13. 2^r.

2.14. $C_r^k C_{n-r}^{s-k}$, если $k = (r+s-t)/2$ — целое неотрицательное число.

2.15. $n-2$.

2.16. $\min\{i, n-i\}$.

2.17. Перестановкой компонент и их инвертированием наборы $\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}$ преобразуются к виду, показанному на рисунке, где тонкие линии обозначают нулевые компоненты, а толстые — единичные. При этом набор $\tilde{\delta}$ необходимо примет вид, указанный на рисунке.



2.18. а) полная, но не базис; б) образует базис; в) не является полной, так как лежит внутри T_1; г) полная, но не базис.

2.19. Это функции $x_1, x_2, \ldots, x_n$.

2.20. Ясно, что $M \subseteq T_0 \cup T_1$. Далее, из $f \in L$ и $f \notin T_0 \cup T_1$ вытекает, что $f \in S$. Следовательно $L \subseteq T_0 \cup T_1 \cup S$. Поэтому искомое число функций есть

$$|P_2^n \cap (T_0 \cup T_1 \cup S)| = |T_0^n \cup T_1^n \cup S^n| =$$

$$= 2|T_0^n| + |S^n| - |T_0^n \cap T_1^n| - 2|T_0^n \cap S^n| + |T_0^n \cap T_1^n \cap S^n| =$$

$$=2\cdot 2^{2^n-1}+2^{2^{n-1}}-2^{2^n-2}-2\cdot 2^{2^{n-1}-1}+2^{2^{n-1}-1}=$$

$$=3\cdot 2^{2^n-2}+2^{2^{n-1}-1}\sim\frac{3}{4}|\mathrm{P}_2^n|,\quad n\to\infty.$$

2.21. Следует из решения предыдущей задачи и теоремы Поста.

2.22. $|\mathrm{P}_2^n\setminus(\mathrm{T}_0\cup\mathrm{T}_1\cup\mathrm{S})|=2^{2^n-2}-2^{2^{n-1}-1}\sim\frac{3}{4}|\mathrm{P}_2^n|,\quad n\to\infty.$

2.23. Следует из решения предыдущей задачи.

2.24. В случае проверки на выполнимость к. н. ф. приводится к виду $(x_n\vee A)(\overline{x}_n\vee B)C$, где A, B и C — к. н. ф. от переменных $x_1,\ldots,x_{n-1}$. Исходная к. н. ф. выполнима в том и только в том случае, если выполнима формула $(A\vee B)C$.

2.25. $f(y_1,y_2,y_3,y_4)=\operatorname{sgn}(2y_1+y_2+y_3+y_4)$.

2.26. Если существуют две нижние единицы, для которых не выполнено хотя бы одно из двух условий задачи, то они могут быть отсечены одним неравенством. Если, напротив, две нижние единицы отсекаются одним неравенством, то из 2-несуммируемости задаваемой этим неравенством пороговой функции следует существование двух нижних единиц, для которых не выполнено хотя бы одно из двух условий задачи.

2.27. Пусть $\mathbf{a}_0=(1,1,\ldots,1)$, $\mathbf{a}_1=(-1,1,\ldots,1)$, …, $\mathbf{a}_n=(-1,-1,\ldots,-1)$, а $\mathbf{b}$ — произвольный n-мерный ±1-вектор. Тогда $(\mathbf{a}_0,\mathbf{b})$ и $(\mathbf{a}_n,\mathbf{b})$ — противоположные чётные числа и $|(\mathbf{a}_{i+1},\mathbf{b})-(\mathbf{a}_i,\mathbf{b})|=2$, $i=0,1,\ldots,n-1$. Поэтому существует такое i, что $(\mathbf{a}_i,\mathbf{b})=0$.

2.28. Число линейно отделимых подмножеств во множестве из K точек общего положения в n-мерном пространстве равно числу конусов, на которые $(n+1)$-мерное пространство разбивается K центрированными гиперплоскостями общего положения, т. е. согласно теореме Штейнера—Шлефли равно $2\sum_{i=0}^{n}C_{K-1}^{i}$. Число же разбиений вдвое меньше, так как каждому разбиению соответствуют два линейно отделимых подмножества.

Оглавление тома 2