

S.I. Bityukov, N.V. Krasnikov

**The use of statistical methods for the search for
new physics at the LHC (in Russian)**

INR RAS, Moscow

IHEP, Protvino

We review ¹ statistical methods used for the search for new physics at LHC.

¹Extended text of lectures given by S.I. Bityukov for students of MIPT and doctoral students of IHEP.

Битюков Сергей Иванович, Красников Николай Валериевич

**Применение статистических методов для поиска
новой физики на Большом Адронном
Коллайдере**

ИЯИ РАН, Москва

ИФВЭ, Протвино

В настоящей работе ² дан обзор статистических методов, используемых при поиске новой физики в экспериментах на Большом адронном коллайдере. Приведены многочисленные примеры, полезные для физиков, занимающихся обработкой данных с детекторов Большого адронного коллайдера. Обзор предназначен для научных работников - как теоретиков, так и экспериментаторов; специалистов в области моделирования физических процессов при столкновениях частиц высоких энергий на современных коллайдерах; всех, кто интересуется извлечением физических результатов из экспериментальных данных.

²Расширенный текст лекций, прочитанный С.И. Битюковым студентам МФТИ и аспирантам ИФВЭ.

Содержание

1	Основные понятия теории вероятностей	8
1.1	Введение. Интуитивное понятие вероятности	8
1.2	Аксиомы Колмогорова	9
1.3	Условные вероятности. Теорема Байеса	9
1.4	Случайные величины. Функции распределения	10
1.5	Свойства функций распределений	11
1.6	Характеристические функции	14
1.7	Основные функции распределения вероятностей	15
1.7.1	Биномиальное распределение	15
1.7.2	Распределение Пуассона	16
1.7.3	Нормальное одномерное распределение (распределение Гаусса)	17
1.7.4	Многомерное нормальное распределение	19
1.7.5	Γ -распределение	22
1.7.6	χ^2 -распределение	24
1.7.7	Логнормальное распределение	26
1.7.8	Равномерное распределение	26
1.8	Закон больших чисел. Центральная предельная теорема	27
1.8.1	Закон больших чисел	27
1.8.2	Центральная предельная теорема	28
1.9	Информация	29
2	Основные задачи и методы математической статистики	31
2.1	Основные задачи статистики	31
2.2	Частотный подход	32
2.2.1	Доверительный интервал и пределы доверия	32
2.2.2	Пределы доверия для параметра распределения Пуассона	33
2.3	Байесовский подход	34
2.4	Метод максимального правдоподобия	37
3	Оценка параметров	39
3.1	Общая постановка задачи	39
3.2	Нормальное распределение	40
3.3	Оценка параметра в распределении Пуассона	45
3.3.1	Частотный подход	46
3.3.2	Метод наибольшего правдоподобия	46
3.3.3	Байесовский метод	48

3.4	Оценка сигнала в распределении Пуассона при ненулевом фоне	49
3.4.1	Частотный подход	49
3.4.2	Байесовский подход	55
3.4.3	Оценка параметров в методе наименьших квадратов	56
3.4.4	Интервал доверия в случае нескольких параметров	57
3.4.5	Асимптотическая нормальность оценки максимального правдоподобия	58
3.5	Ожидаемые пределы	58
4	Учет систематических ошибок	60
4.1	Введение	60
4.2	Оценка параметра фона, исходя из измерений в другой кинематической области	61
4.3	Оценка систематических неопределенностей в методе максимального правдоподобия	62
4.4	Метод усреднения Кузинса-Хайланда	63
5	Проверка гипотез	68
5.1	Введение	68
5.2	Проверка основной гипотезы	69
5.3	Проверка сложной гипотезы	72
5.4	Тест Неймана-Пирсона	74
5.5	Байесовский подход к проверке гипотез	75
6	Комбинирование результатов	78
6.1	Комбинирование двух нормальных распределений	78
6.2	Нормальное распределение. Общий случай	79
6.3	Метод наименьших квадратов	80
6.4	Метод максимального правдоподобия	81
6.5	Комбинирование пределов	83
6.6	Байесовский подход	84
6.7	Комбинирование уровней значимости (комбинирование значений вероятности)	86
6.8	“Look elsewhere” эффект	87
6.9	Тест Колмогорова-Смирнова	88
7	Статистическое программное обеспечение в задачах физики высоких энергий	90
7.1	Обзор основных пакетов	90

7.2	Проект RooStats	91
7.3	Проект BAT	93

Введение

Данный обзор описывает основные статистические методы, используемые при обработке результатов экспериментов CMS [1] и ATLAS [2] в исследованиях на Большом Адронном Коллайдере [3] (БАК). Эти методы реализуются в пакете программ RooStats [4], созданных физиками, участвующими в экспериментах CMS и ATLAS. Пакет программ RooStats базируется на системе ROOT [5], разработанной в ЦЕРНе (Европейском центре ядерных исследований). Основное отличие данного обзора от многочисленных обзоров и монографий по применению статистических методов в экспериментальной физике и других естественнонаучных областях [6, 7, 8, 9, 10] это то, что в обзоре приводятся многочисленные примеры, ориентированные исключительно для нужд физиков, занимающихся обработкой данных с детекторов Большого Адронного Коллайдера. В обзоре почти отсутствуют доказательства каких-либо абстрактных теорем статистики. В каком-то смысле можно сказать, что это “курс молодого бойца”, жаждущего найти проявления новой физики [11] из экспериментальных данных CMS и ATLAS детекторов.

Для чтения этого обзора необходимо знать лишь основы высшей математики (производная, интеграл) в минимальном объеме. Основные понятия теории вероятности кратко приведены в обзоре (Глава 1). Также не требуется каких-либо особо глубоких знаний как в теоретической, так и в экспериментальной физике.

Содержание статьи следующее. В первой главе приведены основные понятия теории вероятностей. Вторая глава на конкретных примерах (ограничения снизу и сверху на параметр) обсуждает основные подходы, используемые в статистике, а именно, частотный и Байесовский подходы и метод максимального правдоподобия. Третья глава описывает методы, которые позволяют извлечь информацию о параметрах сигнала из экспериментальных данных. В качестве статистик используются статистики, полученные из нормального распределения и распределения Пуассона. В четвертой главе приведены методы, позволяющие учесть систематические эффекты, связанные с неточным знанием фонов и сигналов. В пятой главе обсуждаются проблемы проверки гипотез. В нашем случае основной гипотезой является гипотеза о том, что справедлива Стандартная Модель, а альтернативной гипотезой является гипотеза о существовании новой физики вне рамок Стандартной Модели, поиск которой является главной задачей экспериментов на Большом Адронном Коллайдере. В шестой главе обсуждаются методы, применяемые для комбинирования различных экспериментальных данных, в том числе, для комбиниро-

вания данных разных экспериментов, например, экспериментов CMS и ATLAS. В седьмой главе описаны основные возможности программы обработки экспериментальных данных на базе пакета программ RooStats, который в настоящее время является основным программным продуктом, официально рекомендованным коллаборациями CMS и ATLAS для обработки данных.

1 Основные понятия теории вероятностей

1.1 Введение. Интуитивное понятие вероятности

Теория вероятности по сути дела возникла из игры в бросание монеты. Если мы произвольно бросаем монету N раз, то в N_1 случаях выпадет “орел”, а в N_2 случаях выпадет “решка”. Очевидно, что в любом случае выпадет либо “орел”, либо “решка”, то есть справедливо равенство $N = N_1 + N_2$. В случае большого числа бросаний монеты (испытаний) вероятность выпадения “орла” определяется как $P_1 = P(\text{“орел”}) = \frac{N_1}{N}$ и “решки” как $P_2 = P(\text{“решка”}) = \frac{N_2}{N}$. Здесь неявно предполагается, что предел при $N \rightarrow \infty$, который по сути дела и является определением вероятности, существует. Очевидны следующие свойства вероятности:

$$\begin{aligned} P_1 &\geq 0, \\ P_2 &\geq 0, \\ P_1 + P_2 &= 1. \end{aligned} \tag{1}$$

Поскольку стороны “орел” и “решка” практически не отличаются друг от друга, так же часто предполагается (хотя это и не обязательно), что $P(\text{“орел”}) = P(\text{“решка”}) = \frac{1}{2}$.

В качестве более общего случая мы можем представить, что в ящике находится n шаров, на которых нарисованы цифры $1, 2, \dots, n$. Если мы $N \gg n$ раз вытаскиваем один за другим шары из ящика (с возвращением шара сразу же обратно в ящик), то вероятность вынуть шар с номером k ($1 \leq k \leq n$) определяется как $P_k(N) = \frac{n_k}{N}$ (где n_k -количество раз, когда мы вынули шар с номером k). Предполагается, что предел $N \rightarrow \infty$ для всех $P_k(N)$ существует. Это и есть определение вероятности вытаскивать шар k из ящика.

Очевидны следующие свойства вероятностей P_k :

$$\begin{aligned} P_k &\geq 0, \\ \sum_{k=1}^n P_k &= 1. \end{aligned} \tag{2}$$

Нетрудно обобщить вышеприведенные примеры и дать интуитивные определения вероятности на случай чисел $n = 1, 2, \dots, \infty$ и на случай зависимости вероятности (точнее плотности вероятности) от непрерывного

параметра. Интуитивное понятие вероятности можно формализовать на аксиоматическом уровне, что будет показано в следующем параграфе.

1.2 Аксиомы Колмогорова

Понятие вероятности на аксиоматическом уровне можно ввести следующим образом. Пусть у нас имеется некоторое множество Ω . Вероятность это P -вещественная функция, определенная на классе всех подмножеств Ω . Для каждого подмножества A, B, \dots множества Ω функция $P(A)$ удовлетворяет следующим аксиомам (аксиомы Колмогорова [12]):

- $P(A) \geq 0$.
- Для непересекающихся множеств (т.е. $A \cap B = \emptyset$),
 $P(A \cup B) = P(A) + P(B)$.
- $P(\Omega) = 1$.

Множество Ω называется пространством элементарных событий. Подмножества A, B, \dots множества Ω называются событиями.

1.3 Условные вероятности. Теорема Байеса

Условная вероятность $P(A|B)$ наступления события A при условии выполнения события B определяется как

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (3)$$

Из соотношения (3) и из того факта, что $A \cap B = B \cap A$, вытекает теорема Байеса [13]

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (4)$$

Также справедливо очевидное равенство

$$P(B) = \sum_i P(B|A_i)P(A_i) \quad (5)$$

при условии, что $\bigcup_i A_i = \Omega$.

С учетом равенства (5) теорема Байеса (4) может быть сформулирована как

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)}, \quad (6)$$

где $\bigcup_i A_i = \Omega$.

Теорема Байеса является ключевой при формулировании статистики в Байесовском подходе, где она записывается в виде

$$P(\text{theory}|\text{data}) \sim P(\text{data}|\text{theory})P(\text{theory}), \quad (7)$$

где “theory” представляет собой некоторую гипотезу, а “data” представляет собой результат эксперимента. $P(\text{theory})$ является априорной вероятностью для теории, отражающую априорную степень доверия к ней. Функция $P(\text{data}|\text{theory})$ является вероятностью получения данных при заданной теоретической модели “theory” и называется функцией максимального правдоподобия.

События A и B называются независимыми, если вероятность их одновременного осуществления равна произведению вероятностей каждого из них, то есть

$$P(A \cap B) = P(A)P(B). \quad (8)$$

Вспоминая определение (3) условной вероятности, получаем, что для независимых событий A и B

$$\begin{aligned} P(A|B) &= P(A), \\ P(B|A) &= P(B). \end{aligned} \quad (9)$$

1.4 Случайные величины. Функции распределения

Случайной величиной называется функция, определенная на множестве элементарных (событий) исходов. Случайная величина как функция элементарного исхода $\omega \in \Omega$ обычно обозначается как $x(\omega)$. Обычно аргумент ω для краткости опускается.

В случае дискретной величины x_1, x_2, \dots функция $P(X = x_k)$ есть вероятность события $X(\omega) = x_k$.

В случае непересекающихся событий x_k ($x_k \cap x_l = \emptyset$, $k \neq l$), аксиомы Колмогорова сводятся к двум условиям:

$$\begin{aligned} P(x_k) &\geq 0, \\ \sum_k P(x_k) &= 1. \end{aligned} \tag{10}$$

Для непрерывных случайных величин, когда множество всех возможных значений случайных величин заполняют непрерывное или континуальное множество (например, отрезок на прямой, часть плоскости или пространства) вводят понятие распределение вероятности.

Пусть случайная величина x задана при $a \leq x \leq b$. Плотность вероятности $f(x)$ это неотрицательная функция от x , такая, что вероятность наступления события с $x_1 \leq x \leq x_2$ равна $\int_{x_1}^{x_2} f(x)dx = P(x_1 \leq x \leq x_2)$.

Имеет место тривиальное условие нормировки $\int_a^b f(x)dx = 1$.

Обобщение на случай многомерных случайных величин (x_1, x_2, \dots, x_n) очевидно.

При изменении переменных $y = h(x)$ в силу равенства

$$g(y)dy = f(x)dx \tag{11}$$

получаем

$$g(y) = \frac{f(x)}{|h'(x)|}. \tag{12}$$

Кумулятивной функцией распределения $F(y)$ является вероятность того, что $P(x \leq y) = F(y)$. Кумулятивная функция выражается через интеграл от плотности вероятности

$$F(y) = \int_a^y f(x)dx. \tag{13}$$

Заметим, что обычно функция $f(x)$ определяется при $-\infty \leq x \leq \infty$, то есть $b = \infty$, $a = -\infty$.

1.5 Свойства функций распределений

Пусть X это дискретная случайная величина, принимающая значения x_1, x_2, \dots с вероятностями соответственно p_1, p_2, \dots . Математическим ожиданием дискретной случайной величины называется число

$$E(X) = x_1p_1 + x_2p_2 + \dots \tag{14}$$

Для непрерывной случайной величины X обобщение (14) имеет вид

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx. \quad (15)$$

Подчеркнем, что для медленно убывающих функций ³ интеграл (15) может не существовать.

Одним из важных свойств случайной величины является линейность: если a и b два числа, а X и Y две случайные величины, то

$$E(aX + bY) = aE(X) + bE(Y). \quad (16)$$

Свойство (16) непосредственно вытекает из определений (14,15) математических ожиданий для дискретной и непрерывной случайных величин.

Важной характеристикой разброса случайных величин является дисперсия, которая определяется как ⁴

$$D(X) = E(X - E(X))^2. \quad (17)$$

Иными словами, дисперсию можно определить как среднее значение квадрата отклонения случайной величины от своего математического ожидания. Формула (17) эквивалентна формуле

$$D(X) = E(X^2) - (E(X))^2. \quad (18)$$

Справедливо следующее свойство дисперсии:

$$D(cX) = c^2 D(X), \quad (19)$$

где c - некоторое постоянное число (константа).

Часто используют понятие среднеквадратичного отклонения, определяемого как квадратный корень из дисперсии

$$\sigma = \sqrt{D(X)}. \quad (20)$$

Для математического ожидания $E(X)$ справедливо неравенство Маркова

³Типичный пример - распределение Коши с $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$.

⁴В случае медленно убывающих функций, например, $f(x) \sim \frac{1}{1+|x|^3}$, дисперсия не существует, точнее дисперсия равна бесконечности.

$$P(X \geq x) \leq \frac{E(X)}{x}. \quad (21)$$

Для дисперсии $D(X)$ справедливо неравенство Чебышева

$$P(|X - E(X)| \geq x) \leq \frac{D(X)}{x^2}. \quad (22)$$

Неравенства (21,22) следуют из определений $E(X)$, $D(X)$ и свойства неотрицательности плотности распределения вероятностей $f(x)$.

Напомним, что два события A и B являются независимыми, если $P(A \cap B) = P(A) \cdot P(B)$. Это понятие независимости двух событий можно распространить и на случайные величины.

Определение: Случайные величины X и Y называются независимыми, если при любых x и y независимы события $\{\omega : X(\omega) < x\}$ и $\{\omega : Y(\omega) < y\}$.

Можно показать, что если случайные величины независимы, то

$$E(XY) = E(X)E(Y). \quad (23)$$

На языке плотностей вероятностей независимость событий X и Y означает, что плотность вероятности $f(x, y)$ факторизуется:

$$f(x, y) = f_1(x)f_2(y). \quad (24)$$

Свойство (24) можно рассматривать как определение двух независимых случайных величин. Обобщение на случай n независимых величин x_1, x_2, \dots, x_n очевидно, а именно, плотность вероятности $f(x_1, x_2, \dots, x_n)$ должна быть представлена в виде

$$f(x_1, x_2, \dots, x_n) = \prod_{k=1}^n f_k(x_k). \quad (25)$$

Степень зависимости случайных величин характеризуется понятием ковариации случайных величин X и Y . Ковариация определяется как

$$\text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))]. \quad (26)$$

Определение (26) можно переписать в виде

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y). \quad (27)$$

Если $cov(X, Y) = 0$, то говорят, что случайные величины X и Y являются некоррелированными. Легко видеть, что для независимых случайных величин X и Y $cov(X, Y) = 0$. Однако обратное утверждение неверно.

Для дисперсии от суммы двух случайных величин справедлива формула

$$D(X + Y) = D(X) + D(Y) + 2cov(X, Y). \quad (28)$$

Отсюда следует, что для независимых величин дисперсия суммы равна сумме дисперсий.

Часто используют коэффициент корреляции $\rho(x, y)$ двух случайных величин X и Y , определяемый как

$$\rho(X, Y) = \frac{cov(X, Y)}{\sqrt{D(X)D(Y)}}. \quad (29)$$

Можно показать, что

$$|\rho(X, Y)| \leq 1. \quad (30)$$

Более того, если $|\rho(X, Y)| = 1$, то случайные величины X и Y связаны линейной зависимостью, то есть существуют числа a и b такие, что

$$P(Y = aX + b) = 1. \quad (31)$$

При этом если $\rho(X, Y) = 1$, то $a > 0$, и, если $\rho(X, Y) = -1$, то $a < 0$. Можно сказать, что величина коэффициента корреляции характеризует степень линейной зависимости.

1.6 Характеристические функции

Для случайной переменной X с плотностью вероятности $f(x)$ характеристическая функция определяется как

$$\Phi_X(t) = E(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} f(x) dx. \quad (32)$$

Для дискретной случайной величины X с конечным числом значений

$$\Phi_X(t) = \sum_k p_k e^{itx_k}. \quad (33)$$

Характеристическая функция $\Phi_X(t)$ полностью определяет плотность вероятности случайной величины, а именно,

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_X(t) e^{-itx} dt. \quad (34)$$

Для функции $\Phi_X(t)$ справедливы соотношения

$$\Phi_X(0) = 1, \quad (35)$$

$$|\Phi_X(t)| \leq 1, \quad (36)$$

$$\Phi_{aX+b}(t) = e^{ibt} \Phi_X(at). \quad (37)$$

Здесь a и b – константы.

Если X и Y независимые случайные переменные с характеристическими функциями $\Phi_X(t)$ и $\Phi_Y(t)$, то справедливо равенство

$$\Phi_{X+Y}(t) = \Phi_X(t) \cdot \Phi_Y(t). \quad (38)$$

Обобщение равенства (38) на случай N независимых переменных имеет вид

$$\Phi_{X_1+\dots+X_N}(t) = \prod_{i=1}^N \Phi_{X_i}(t). \quad (39)$$

1.7 Основные функции распределения вероятностей

В этом разделе мы приводим основные функции распределения вероятностей, наиболее часто используемые в физике высоких энергий.

1.7.1 Биномиальное распределение

Для биномиального распределения вероятность наступления r событий при N испытаниях (r – целое неотрицательное число, $1 \leq r \leq N$) равна

$$P(r) = \frac{N!}{r!(N-r)!} p^r (1-p)^{N-r}, \quad (40)$$

где $0 \leq p \leq 1$ ожидаемая вероятность появления события в одном испытании. Среднее значение при этом равно

$$E(r) = Np, \quad (41)$$

а дисперсия равна

$$D(r) = Np(1 - p). \quad (42)$$

Биномиальное распределение представляет решение следующей задачи. Пусть имеется общая (или как еще говорят генеральная) совокупность, состоящая из M некоторых объектов (например, это могут быть события в экспериментальной установке), из которых $R = pM$ объектов обладают некоторым признаком C , а остальные $M - R = (1 - p)M$ не обладают им, так что $pM + (1 - p)M = M$. Пусть из этой общей совокупности сделана выборка путем независимого случайного выбора одного за другим N объектов таким образом, чтобы не нарушилось соотношение между объектами с признаком C и без признака C . Это возможно в двух случаях: либо общая совокупность бесконечна, либо выбранные объекты возвращаются в общую совокупность перед каждым новым выбором. В таком случае вероятность того, что из общего числа N объектов выборки r объектов будут обладать признаком C будет определяться формулой (40).

Выбор объектов с признаком C или без признака C из общей совокупности называют испытаниями Бернулли. При этом выбор объекта с признаком C называют событием или успехом.

На Рис. 1 показаны распределения вероятностей для биномиального распределения в случае $N = 40$ испытаний с разными значениями параметра p : слева $p = 0.4$, справа $p = 0.8$.

1.7.2 Распределение Пуассона

Для распределения Пуассона функция распределения вероятностей имеет вид

$$P(r|\mu) = \frac{\mu^r e^{-\mu}}{r!}, \quad (43)$$

где r целое неотрицательное число, а параметр $\mu \geq 0$ – действительное число. Среднее значение и дисперсия равны

$$E(r) = \mu, \quad (44)$$

$$D(r) = \mu. \quad (45)$$

Распределение Пуассона можно рассматривать как предельный случай биномиального распределения при выполнении следующего условия. Пусть независимые испытания производятся сериями длины n (n -я

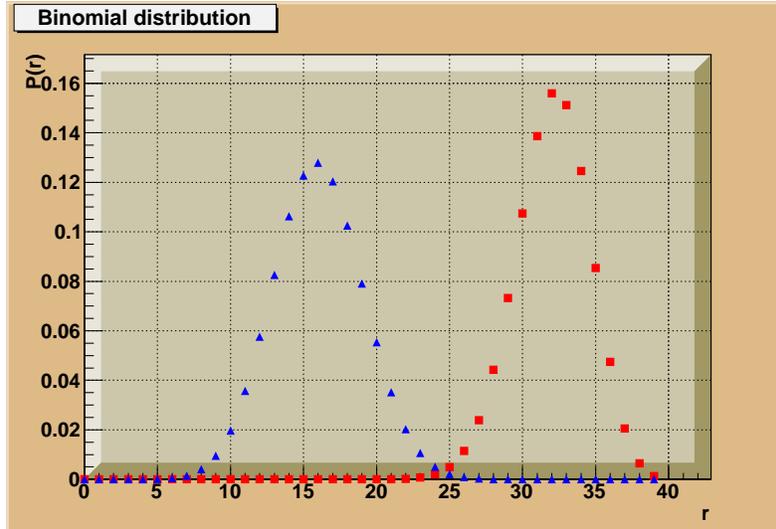


Рис. 1: Распределения вероятностей для биномиального распределения в случае $N = 40$ испытаний с разными значениями параметра p : слева $p = 0.4$, справа $p = 0.8$.

серия, следовательно, состоит из n независимых испытаний), а вероятность $p^{(n)}$ некоторого события для всех испытаний n -й серии одна и та же, но с увеличением n стремится к нулю так, что произведение $np^{(n)}$ сохраняет постоянное значение μ :

$$\lim_{n \rightarrow \infty} np^{(n)} = \mu, \quad (46)$$

причем очевидно, что $\mu > 0$. При этом условии в пределе $n \rightarrow \infty$ вероятность $P(r|\mu)$ того, что рассматриваемое событие в бесконечно длинной серии испытаний появится ровно r раз описывается формулой (43).

На Рис. 2 показаны распределения вероятностей для распределения Пуассона с разными значениями параметров: слева $\mu = 10$, справа $\mu = 20$.

1.7.3 Нормальное одномерное распределение (распределение Гаусса)

Пусть дана случайная величина X , подчиняющаяся нормальному распределению с параметрами μ и σ . Параметры являются действительными числами: $-\infty \leq \mu \leq \infty$ и $0 \leq \sigma \leq \infty$.

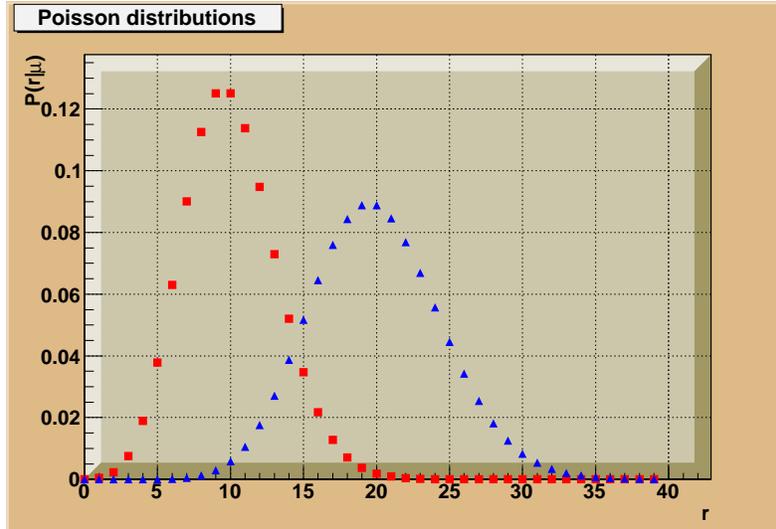


Рис. 2: Распределения вероятностей для распределения Пуассона с разными значениями параметров: слева $\mu = 10$, справа $\mu = 20$.

Плотность вероятности для одномерного нормального распределения равна

$$f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}, \quad (47)$$

где x – действительное число, $-\infty \leq x \leq \infty$. Среднее значение и дисперсия равны

$$E(X) = \mu, \quad (48)$$

$$D(X) = \sigma^2. \quad (49)$$

Очень важное свойство нормального распределения состоит в том, что, если две случайные величины X и Y подчиняются нормальному распределению со средними μ_X и μ_Y и дисперсиями σ_X и σ_Y соответственно, то случайная величина $Z = X + Y$ распределена согласно нормальному распределению со средним $\mu_Z = \mu_X + \mu_Y$ и дисперсией $\sigma_Z = \sqrt{\sigma_X^2 + \sigma_Y^2}$. Доказательство этого факта заключается во взятии интеграла в формуле

$$F(z|\mu_X, \mu_Y, \sigma_X, \sigma_Y) = \int dx dy \delta(z - x - y) N(x|\mu_X, \sigma_X) N(y|\mu_Y, \sigma_Y).$$

Отсюда, в частности, следует, что если случайные величины X_1, X_2, \dots, X_N подчиняются нормальному распределению с универсальными средним μ и дисперсией σ , то случайная величина $X(N) = \frac{1}{N} \sum_{k=1}^N X_k$ подчиняется нормальному распределению со средним $\mu_X = \mu$ и дисперсией $\sigma_X^2 = \frac{1}{N} \sigma^2$. В пределе $N \rightarrow \infty$ случайная величина $X(N)$ с вероятностью равной 1 принимает значение μ .

Характеристическая функция для нормального распределения равна

$$\Phi(t) = e^{it\mu - \frac{1}{2}t^2\sigma^2}. \quad (50)$$

Нормальное распределение также можно получить путем предельного перехода из биномиального распределения. Так, если X – число событий (успехов) в n испытаниях Бернулли, а вероятность появления события в одном испытании остается той же самой, то какими бы не были действительные числа a и b ($a < b$),

$$P(a \leq \frac{X - np}{\sqrt{np(1-p)}} \leq b) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

Это утверждение называется теоремой Муавра-Лапласа.

Весьма важной является также теорема Ляпунова, которая утверждает, что если значения независимых случайных величин будут малы в сравнении с их суммой, то, при неограниченном возрастании числа этих величин, распределение их суммы становится приближенно нормальным. Теорема работает при весьма общих условиях для любых распределений. Это позволяет при исследовании случайной величины, являющейся суммой многих случайных величин, влияние которых на саму сумму незначительно, заранее предполагать, что распределение ошибок при измерении этой случайной величины будет нормальным.

На Рис. 3 показаны плотности вероятности для нормального распределения с разными значениями параметров: слева $\mu = 10, \sigma = 5$, справа $\mu = 20, \sigma = 10$.

1.7.4 Многомерное нормальное распределение

Пусть дана k -мерная случайная величина \vec{X} , подчиняющаяся многомерному нормальному распределению с параметрами $\vec{\mu}$ (k -мерный вектор

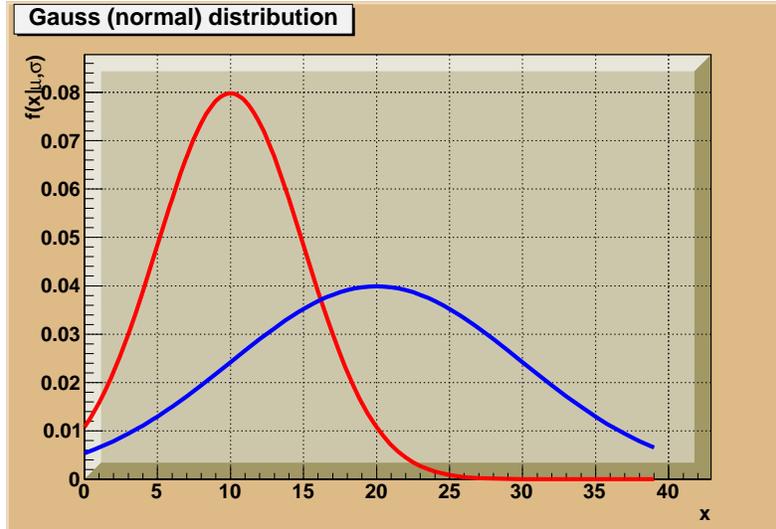


Рис. 3: Распределения плотности вероятности для нормального распределения с разными значениями параметров: слева $\mu = 10$, $\sigma = 5$, справа $\mu = 20$, $\sigma = 10$.

действительных чисел) и $[V]$ (матрица размерности $[k \times k]$ с положительными собственными числами).

Плотность вероятности равна

$$f(X|\vec{\mu}, [V]) = \frac{1}{(2\pi)^{\frac{k}{2}} V^{\frac{1}{2}}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T [V]^{-1}(\vec{x}-\vec{\mu})}, \quad (51)$$

где переменная $\vec{x} = (x_1, x_2, \dots, x_k)$ – k -мерный вещественный вектор, $V = \det[V]$.

Среднее значение равно

$$E(\vec{X}) = \vec{\mu}. \quad (52)$$

Ковариация определяется как

$$\text{cov}(X_i, X_j) \equiv E(X_i X_j) - E(X_i)E(X_j), \quad i, j = 1, k. \quad (53)$$

Корреляция определяется как

$$\rho_{ij} = \frac{\text{cov}(X_i, X_j)}{\sigma_{X_i} \sigma_{X_j}}, \quad (54)$$

где $\sigma_{X_i}^2 = E((X_i - E(X_i))^2)$. Нетрудно доказать, используя неравенство Коши-Буняковского, что $-1 \leq \rho_{ij} \leq 1$.

Дисперсия и ковариация равны, соответственно,

$$D(X_i) = V_{ii}, \quad i = 1, k, \quad (55)$$

$$\text{cov}(X_i, X_j) = V_{ij}, \quad i, j = 1, k. \quad (56)$$

Рассмотрим в качестве часто встречающегося примера двумерное нормальное распределение ⁵.

Путем замены координат в плоскости (X, Y)

$$X' = X \cos \varphi - Y \sin \varphi$$

$$Y' = X \sin \varphi + Y \cos \varphi$$

можно выбрать новую систему координат, в которой функция

$$(X, Y)^T [V]^{-1} (X, Y) = \frac{X'^2}{\sigma_{X'}^2} + \frac{Y'^2}{\sigma_{Y'}^2}$$

диагональна и плотность нормального распределения $N(X', Y' | \dots)$ принимает вид

$$f(x', y') = \frac{1}{2\pi\sigma_{X'}\sigma_{Y'}} e^{-\frac{1}{2}\left(\frac{x'^2}{\sigma_{X'}^2} + \frac{y'^2}{\sigma_{Y'}^2}\right)}.$$

Вследствие того, что вращение сохраняет детерминант матрицы $[V]^{-1}$ имеем равенство

$$\frac{1}{\sigma_{X'}^2 \sigma_{Y'}^2} = \frac{1}{\sigma_X^2 \sigma_Y^2 (1 - \rho_{XY}^2)}.$$

Матрица $[V]$ и коэффициент корреляции ρ_{XY} в новых координатах

$$[V] = \begin{pmatrix} \sigma_X^2 & \rho_{XY} \sigma_X \sigma_Y \\ \rho_{XY} \sigma_X \sigma_Y & \sigma_Y^2 \end{pmatrix} =$$

$$\begin{pmatrix} \cos^2 \varphi \sigma_{X'}^2 + \sin^2 \varphi \sigma_{Y'}^2 & \sin \varphi \cos \varphi (\sigma_{Y'}^2 - \sigma_{X'}^2) \\ \sin \varphi \cos \varphi (\sigma_{Y'}^2 - \sigma_{X'}^2) & \sin^2 \varphi \sigma_{X'}^2 + \cos^2 \varphi \sigma_{Y'}^2 \end{pmatrix},$$

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\sin 2\varphi (\sigma_{Y'}^2 - \sigma_{X'}^2)}{\sqrt{\sin 2\varphi (\sigma_{Y'}^4 - \sigma_{X'}^4) + 2\sigma_{Y'}^2 \sigma_{X'}^2}},$$

$$\tan 2\varphi = \frac{2\rho_{XY} \sigma_X \sigma_Y}{\sigma_{Y'}^2 - \sigma_{X'}^2},$$

⁵Здесь для простоты мы рассматриваем случай $\mu_X = \mu_Y = 0$. Обобщение для общего случая очевидно.

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}} e^{-\frac{1}{2(1-\rho_{XY}^2)}\left(\frac{x^2}{\sigma_X^2} + \frac{y^2}{\sigma_Y^2} - \frac{2xy\rho_{XY}}{\sigma_X\sigma_Y}\right)}.$$

При проецировании функции распределения на оси X и Y получаем

$$f_X(x) \equiv \int_{-\infty}^{\infty} f(x, y)dy = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{x^2}{2\sigma_X^2}},$$

$$f_Y(y) \equiv \int_{-\infty}^{\infty} f(x, y)dx = \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{y^2}{2\sigma_Y^2}}.$$

Характеристическая функция равна

$$\Phi(\vec{t}) = e^{i\vec{t}\vec{\mu} - \frac{1}{2}\vec{t}^T[V]\vec{t}}. \quad (57)$$

На Рис. 4 показана плотность вероятности для двумерного нормального распределения со следующими значениями параметров: $\mu_1 = 20$, $\sigma_1 = 10$, $\mu_2 = 10$, $\sigma_2 = 5$, $\rho_{12} = 0$.

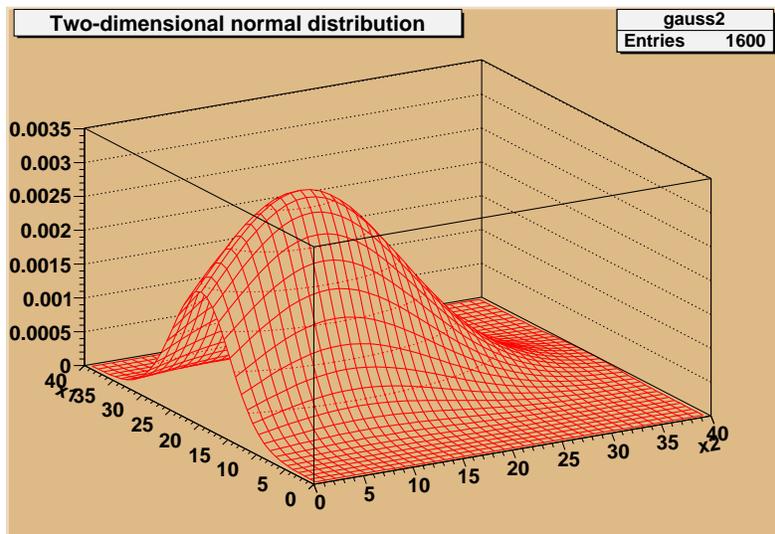


Рис. 4: Распределения плотности вероятности для двумерного нормального распределения со следующими значениями параметров: $\mu_1 = 20$, $\sigma_1 = 10$, $\mu_2 = 10$, $\sigma_2 = 5$, $\rho_{12} = 0$.

1.7.5 Г-распределение

Г-распределение относится к числу основных типов распределений. Напомним, что Г-функция

$$\Gamma(b) = \int_0^{\infty} x^{b-1} e^{-x} dx = (b-1)! \quad (58)$$

представляет собой интеграл, равный факториалу какого угодно положительного числа b . Этот факт можно реализовать в виде функционального уравнения Γ -функции, представляющего основное свойство Γ -функции,

$$\Gamma(b+1) = b\Gamma(b).$$

Распределение непрерывной случайной величины X , зависящей от одного параметра b с плотностью вероятности

$$f(x|b) = \frac{x^{b-1} e^{-x}}{\Gamma(b)}$$

называется Γ -распределением. Кумулятивная функция распределения есть

$$F(x|b) = \frac{1}{\Gamma(b)} \int_0^x x'^{b-1} e^{-x'} dx'.$$

Математическое ожидание и дисперсия случайной величины совпадают и равны значению параметра b .

Также Γ -распределением $\Gamma(a, b)$ называется распределение с двумя параметрами a и b и с плотностью вероятности

$$f(x|a, b) = \frac{a(ax)^{b-1} e^{-ax}}{\Gamma(b)}, \quad (59)$$

где a и b положительные действительные числа. В этом случае среднее значение и дисперсия равны

$$E(X) = \frac{b}{a}, \quad (60)$$

$$D(X) = \frac{b}{a^2}. \quad (61)$$

Характеристическая функция равна

$$\Phi(t) = \left(1 - \frac{it}{a}\right)^{-b}. \quad (62)$$

Γ -распределение тесно связано с нормальным распределением. Параметр b определяет форму распределения, а параметр a является параметром масштаба.

На Рис. 5 показаны плотности вероятности для Γ -распределения с разными значениями параметров: слева $a = 1.2$, $b = 5$, справа $a = 8$, $\sigma = 2$.

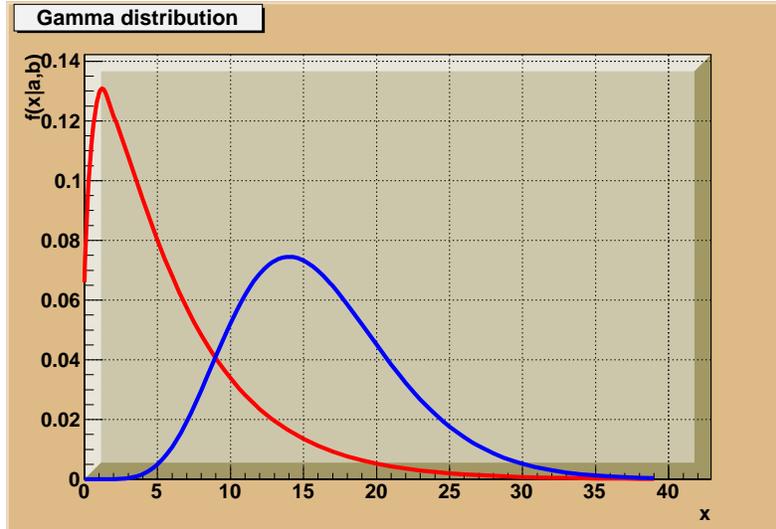


Рис. 5: Распределения плотности вероятности для Γ -распределения с разными значениями параметров: слева $a = 1.2$, $b = 5$, справа $a = 8$, $b = 2$.

1.7.6 χ^2 -распределение

Для χ^2 -распределения плотность вероятности равна

$$f(x|N) = \frac{\frac{1}{2} \left(\frac{x}{2}\right)^{\frac{N}{2}-1}}{\Gamma\left(\frac{N}{2}\right)} e^{-\frac{x}{2}}, \quad (63)$$

где $0 \leq x \leq \infty$, а N -целое положительное число.

Среднее значение и дисперсия равны

$$E(X) = N, \quad (64)$$

$$D(X) = 2N. \quad (65)$$

Характеристическая функция равна

$$\Phi(t) = (1 - 2it)^{-\frac{N}{2}}. \quad (66)$$

χ^2 -распределение связано с нормальным распределением. А именно, пусть X_1, X_2, \dots, X_N являются независимыми случайными величинами,

подчиняющимися стандартному нормальному распределению (то есть $\mu_n = 0$, $D(X_n) = 1$, $n = 1, \dots, N$). Тогда сумма квадратов

$$\chi_N^2 = \sum_{i=1}^N x_i^2 \quad (67)$$

распределена по χ^2 -распределению с N степенями свободы. χ^2 -распределение возникло с развитием статистики для описания поведения функций от наблюдаемых данных и является частным случаем Γ -распределения

$$\chi_N^2 = \Gamma\left(\frac{1}{2}, \frac{N}{2}\right).$$

На Рис. 6 показаны плотности вероятности для χ^2 -распределения с разными значениями параметров: слева $N = 10$, справа $N = 20$.

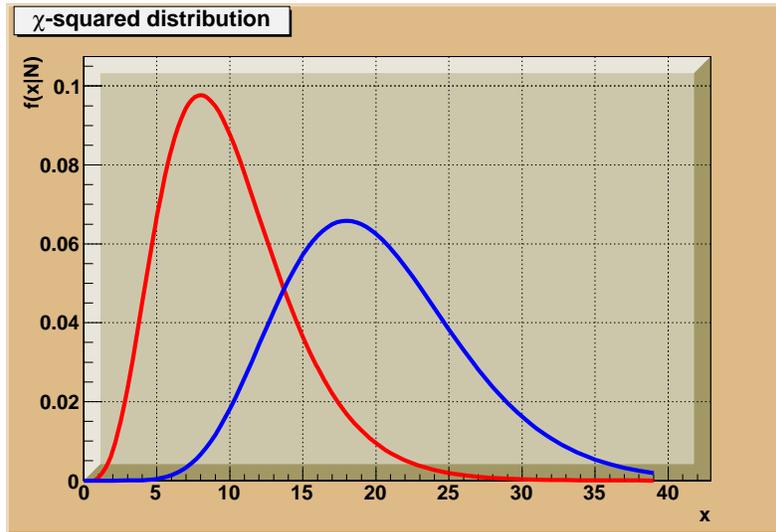


Рис. 6: Распределения плотности вероятности для χ^2 -распределения с разными значениями параметров: слева $N = 10$, справа $N = 20$.

Функция $\chi_N = \sqrt{\chi_N^2}$ имеет $\chi(N)$ -распределение с функцией плотности

$$f(x|N) = \frac{\left(\frac{1}{2}\right)^{\frac{N}{2}-1} x^{N-1}}{\Gamma\left(\frac{N}{2}\right)} e^{-\frac{1}{2}x^2}.$$

В асимптотическом пределе $\chi^2(N)$ - и $\chi(N)$ -распределения стремятся к нормальному, причем эти распределения с хорошей точностью можно считать нормальными при $n > 30$.

Можно также показать, что при больших значениях N величины $S_N = \frac{\chi_N^2 - N}{\sqrt{2N}}$ и $T_N = \sqrt{2\chi_N^2} - \sqrt{2N - 1}$ распределены по стандартному нормальному закону $N(0, 1)$.

1.7.7 Логнормальное распределение

Плотность вероятности логарифмически нормального распределения (логнормального) равна

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{x} e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2}, \quad (68)$$

где μ -действительное число, а σ -положительное действительное число.

Среднее значение и дисперсия равны

$$E(X) = e^{\mu - \frac{1}{2}\sigma^2}, \quad (69)$$

$$D(X) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1). \quad (70)$$

Логнормальное распределение описывает случайную переменную, логарифм которой подчиняется нормальному распределению.

На Рис. 7 показаны плотности вероятности для логнормального распределения с разными значениями параметров: слева $\mu = 5$, $\sigma = 2.5$, справа $\mu = 4$, $\sigma = 0.9$.

1.7.8 Равномерное распределение

Плотность вероятности равна

$$f(x|a, b) = \frac{1}{b - a}, \quad (71)$$

где x – действительная переменная, удовлетворяющая неравенству $a \leq x \leq b$, a и b – действительные числа.

Среднее значение и дисперсия равны

$$E(X) = \frac{a + b}{2}, \quad (72)$$

$$D(X) = \frac{(b - a)^2}{12}. \quad (73)$$

Характеристическая функция равна

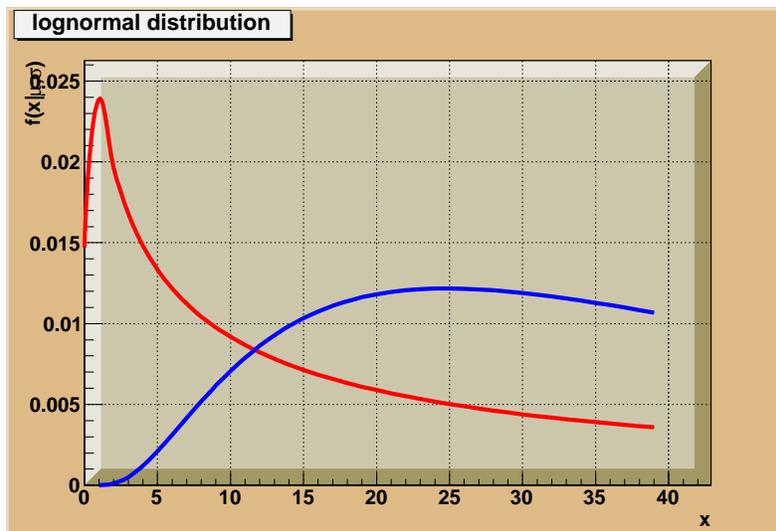


Рис. 7: Распределения плотности вероятности для логнормального распределения с разными значениями параметров: слева $\mu = 5$, $\sigma = 2.5$, справа $\mu = 4$, $\sigma = 0.9$.

$$\Phi(\vec{t}) = \frac{e^{itb} - e^{ita}}{it(b-a)}. \quad (74)$$

На Рис. 8 показаны плотности вероятности для равномерного распределения с разными значениями параметров: слева $a = 5$, $b = 10$, справа $a = 15$, $b = 35$.

1.8 Закон больших чисел. Центральная предельная теорема

1.8.1 Закон больших чисел

Пусть X_1, X_2, \dots, X_N - независимые одинаково распределенные случайные величины и пусть $E(X) = \mu$. Тогда с вероятностью 1 выполняется соотношение

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N X_k = \mu. \quad (75)$$

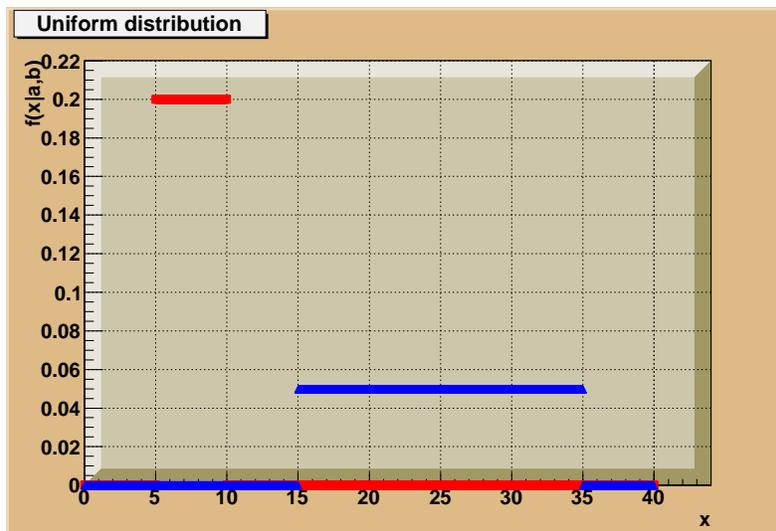


Рис. 8: Распределения плотности вероятности для равномерного распределения с разными значениями параметров: слева $a = 5$, $b = 10$, справа $a = 15$, $b = 35$.

Закон больших чисел (75) следует из неравенства Чебышева (22) и из того факта, что для независимых величин X_1, \dots, X_N с одинаковым средним значением и дисперсией

$$D\left(\frac{X_1 + \dots + X_N}{N}\right) = \frac{D(X_1)}{N}. \quad (76)$$

1.8.2 Центральная предельная теорема

Центральная предельная теорема утверждает, что сумма N независимых случайных величин X_k стремится в пределе $N \rightarrow \infty$ к нормальному распределению, а именно

$$\frac{\sum_{k=1}^N X_k - \sum_{k=1}^N \mu_k}{\sqrt{\sum_{k=1}^N \sigma_k^2}} \rightarrow N(0, 1). \quad (77)$$

Здесь μ_k и σ_k^2 – средние значения и дисперсии случайных величин X_k , а

$N(0, 1)$ стандартное нормальное распределение с плотностью распределения

$$f(x|\mu = 0, \sigma = 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \quad (78)$$

Доказательство теоремы можно найти, например, в монографии [10]. Центральная предельная теорема по сути дела иллюстрирует тот факт, что нормальное распределение явно выделено из всех других распределений.

1.9 Информация

Рассмотрим N независимых измерений случайной величины X , зависящей от вектора параметров $\vec{\theta}$. Их можно рассматривать как случайный вектор с N случайными одинаково распределенными компонентами X_k , $k = 1, 2, \dots, N$, то есть как $\vec{X} = (X_1, X_2, \dots, X_N)$. Множество всех допустимых значений \vec{X} обозначим как $\Omega_{\vec{\theta}}$.

В силу независимости X_k плотность вероятности \vec{X} равна

$$P(\vec{X}|\vec{\theta}) = P(X_1, X_2, \dots, X_n|\vec{\theta}) = \prod_{k=1}^N f_k(x, \vec{\theta}), \quad (79)$$

где $f_k(x, \vec{\theta})$ – плотность вероятности случайной величины X_k , а $\vec{\theta}$ – параметры, описывающие распределение. В случае, когда переменная \vec{X} заменяется на наблюдаемые данные $\vec{x} = (x_1, x_2, \dots, x_N)$ функция P уже не является функцией распределения случайных величин, а называется функцией правдоподобия, которая зависит от параметров распределения $\vec{\theta}$

$$L(\vec{x}|\vec{\theta}) = P(\vec{x}|\vec{\theta}). \quad (80)$$

Любая функция $T = T(x_1, x_2, \dots, x_N)$ от измеренных величин называется статистикой.

Информация по Фишеру [14] определяется как

$$I_x(\theta) = E\left[\left(\frac{\partial \ln L(x|\theta)}{\partial \theta}\right)^2\right] = \int_{\Omega_{\vec{\theta}}} \left(\frac{\partial \ln L(x|\theta)}{\partial \theta}\right)^2 L(x|\theta) dx. \quad (81)$$

В случае, когда $\vec{\theta} = (\theta_1, \theta_2, \dots, \theta_N)$ - k -мерный вектор, обобщение формулы (81) есть формула

$$\begin{aligned}
[I_x(\vec{\theta})]_{ij} &= E\left[\frac{\partial \ln L(x|\vec{\theta})}{\partial \theta_i} \cdot \frac{\partial \ln L(x|\vec{\theta})}{\partial \theta_j}\right] \\
&= \int_{\Omega_{\vec{\theta}}} \left[\frac{\partial \ln L(x|\vec{\theta})}{\partial \theta_i} \cdot \frac{\partial \ln L(x|\vec{\theta})}{\partial \theta_j}\right] L(x|\vec{\theta}) dx.
\end{aligned} \tag{82}$$

В случае, когда случайная величина X распределена по нормальному закону с неизвестным средним μ и известной дисперсией σ^2 , информация для одного наблюдения x равна

$$\begin{aligned}
I_1(\mu, \sigma) &= - \int_{-\infty}^{\infty} \frac{\partial^2}{\partial \mu^2} \left(-\frac{x-\mu}{2\sigma^2} - \ln \sigma\sqrt{2\pi}\right) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx = \\
&= \int_{-\infty}^{\infty} \frac{1}{\sigma^2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx = \frac{1}{\sigma^2}.
\end{aligned} \tag{83}$$

В случае N независимых измерений x_1, x_2, \dots, x_N случайной величины X , подчиняющейся нормальному распределению, информация равна

$$\begin{aligned}
I_N(\mu, \sigma) &= - \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{\partial^2}{\partial \mu^2} \left(\sum_{i=1}^N \left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) - \ln \sigma\sqrt{2\pi}\right) \times \\
&= \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x_i - \mu)^2}{\sigma^2}} dx_i = \frac{N}{\sigma^2} = NI_1(\mu, \sigma).
\end{aligned} \tag{84}$$

В дискретном случае, когда у нас случайная величина X принимает значения $1, 2, \dots, N$ с вероятностями p_1, p_2, \dots, p_N соответственно, информация определяется как

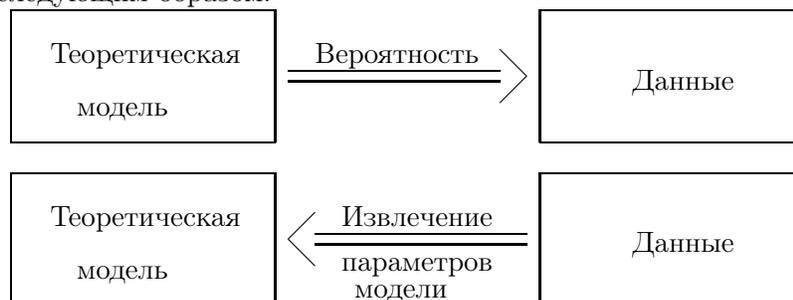
$$H(p_1 \dots p_N) = -k \sum_{k=1}^N p_i \ln p_i. \tag{85}$$

где k -некоторая нормировочная константа, а $\sum_i p_i = 1$, $p_i \geq 0$. Можно показать, что информация (85) максимальна при $p_1 = p_2 = \dots = p_N = \frac{1}{N}$.

2 Основные задачи и методы математической статистики

2.1 Основные задачи статистики

Если основной задачей теории вероятностей является построение модели случайных явлений, позволяющих выразить вероятности более сложных явлений через вероятности простых, то задачи математической статистики являются в некотором смысле обратными по отношению к теории вероятностей. И состоят они в выработке процедур и алгоритмов, позволяющих по экспериментальным данным получить оценки параметров функций распределений вероятностей и, тем самым, оценить параметры теоретической модели. Итак, графически мы можем проиллюстрировать это следующим образом.



Можно сказать, что основными задачами статистики являются.

1. Оценка параметров модели из имеющихся экспериментальных данных.
2. Построение интервалов доверия – нахождение областей параметров модели, не противоречащих экспериментальным данным.
3. Проверка гипотез – удовлетворяет ли та или иная гипотеза экспериментальным данным и какая из рассматриваемых гипотез удовлетворяет экспериментальным данным лучше.
4. Качество фита – насколько хорошо модель соответствует экспериментальным данным.

На сегодняшний день в математической статистике используются в основном три метода, позволяющих решить обратную задачу – извлечение из экспериментальных данных параметров распределения:

- частотный подход,

- метод максимального правдоподобия,
- Байесовский метод.

Заметим, что частотный подход близок к методу максимального правдоподобия, но не совпадает с ним. Численные оценки, полученные с помощью этих трех методов как правило различаются. В следующих разделах этой главы мы описываем эти три метода на примере задачи нахождения верхнего предела на параметр λ распределения Пуассона $P(n|\lambda)$.

2.2 Частотный подход

2.2.1 Доверительный интервал и пределы доверия

Часто на основании измерений случайной переменной делаются выводы о возможных значениях параметров распределений, описывающих поведение этой переменной. Так как обычно параметр не является случайной величиной, то мы не можем говорить о вероятностных характеристиках параметра. В этом случае можно говорить о степени доверия к оценке величины параметра, полученной из результатов измерений. Для представления оценок параметров были введены понятия доверительный интервал и пределы доверия ⁶. Доверительное оценивание в частотной интерпретации часто связывают с Е. Нейманом [16].

Доверительный интервал - это интервал, построенный с помощью случайной выборки из распределения с неизвестным параметром, такой, что он содержит данный параметр с заданной вероятностью (уровнем доверия) $1 - \alpha$, где величина $\alpha \in [0, 1]$ обычно называется уровнем значимости.

Определение: Пусть $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n$ есть выборка из распределения $P(\theta)$, где θ - неизвестный параметр. Пусть также задан уровень доверия $1 - \alpha \in [0, 1]$. Тогда случайный интервал $[L, U]$, где $L = L(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n)$, $U = U(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n)$ есть некоторые статистики имеющейся выборки, такой, что $P(L \leq \theta \leq U) = 1 - \alpha$, называется $(1 - \alpha)$ -доверительным интервалом для параметра θ . При одностороннем построении доверительного интервала для параметра в качестве случайной величины выступает только одна из границ интервала. В этом случае, если этой границей является левая граница интервала L , то говорят о нижнем пределе с

⁶Существует подход, позволяющий расширить эти понятия до понятий доверительная плотность и доверительное распределение. Более подробно с данным направлением исследований можно познакомиться в обзоре [15].

уровнем доверия $(1 - \alpha)$. Если же в качестве такой границы выступает правая граница U , то говорят об верхнем $(1 - \alpha)$ -пределе.

2.2.2 Пределы доверия для параметра распределения Пуассона

Пусть мы измеряем количество событий в эксперименте, причем события распределены по закону Пуассона $P(n|\lambda) = \frac{\lambda^n}{n!} e^{-\lambda}$. Предположим мы наблюдаем n_{obs} событий и наша цель получить верхний предел $\lambda \leq \lambda_{up}$ на параметр λ с точностью $(1 - \alpha)$. Обычно на практике α выбирается равной 0.05 (95% уровень достоверности). В частотном подходе требуется, чтобы

$$P(n \leq n_{obs}|\lambda) \equiv \sum_{n=0}^{n_{obs}} P(n|\lambda) \geq \alpha. \quad (86)$$

Иными словами, вероятность наблюдения количества событий меньшей или равной n_{obs} должна быть не меньше, чем α . Из соотношения (86) можно получить ограничение сверху на параметр λ в распределении Пуассона

$$\lambda \leq \lambda_{up}(n_{obs}, \alpha). \quad (87)$$

Верхний предел зависит как от n_{obs} , так и от α . Так, например,

$$\lambda_{up}(n_{obs} = 0, \alpha = 0.05) = 3.0, \quad (88)$$

$$\lambda_{up}(n_{obs} = 0, \alpha = 0.01) = 4.6, \quad (89)$$

$$\lambda_{up}(n_{obs} = 2, \alpha = 0.05) = 6.3, \quad (90)$$

$$\lambda_{up}(n_{obs} = 2, \alpha = 0.01) = 8.4, \quad (91)$$

Задача определения нижнего предела λ_{low} на уровне $1 - \alpha$ достоверности сводится к решению неравенства

$$P(n \geq n_{obs}|\lambda) = \sum_{n=n_{obs}}^{\infty} P(n|\lambda) \geq \alpha.$$

В Таб. 1 представлены верхние λ_{up} и нижние λ_{low} пределы для $n_{obs} \leq 10$ и $\alpha = 0.1, \alpha = 0.05, \alpha = 0.01$ [17].

Таблица 1: Значения верхних λ_{up} и нижних λ_{low} пределов для $n_{obs} \leq 10$ и $\alpha = 0.05$, $\alpha = 0.1$.

n_{obs}	λ_{low}	λ_{up}	λ_{low}	λ_{up}	λ_{low}	λ_{up}
	$1 - \alpha = 0.9$		$1 - \alpha = 0.95$		$1 - \alpha = 0.99$	
0	—	2.30	—	3.00	—	4.60
1	0.105	3.89	0.051	4.74	0.01	6.63
2	0.532	5.32	0.355	6.30	0.15	8.40
3	1.10	6.68	0.818	7.75	0.44	10.04
4	1.74	7.99	1.37	9.15	0.82	11.60
5	2.43	9.27	1.97	10.51	1.28	13.10
6	3.15	10.53	2.61	11.84	1.78	14.57
7	3.89	11.77	3.29	13.15	2.33	15.99
8	4.66	12.99	3.98	14.43	2.90	17.40
9	5.43	14.21	4.70	15.71	3.51	18.78
10	6.22	15.41	5.43	16.96	4.13	20.14

Подчеркнем, что иногда этот подход к определению верхнего предела критикуется на том основании, что в формуле (86) содержится суммирование по ненаблюдаемым вероятностям $P(n_{obs} - 1, \lambda)$, $P(n_{obs} - 2, \lambda), \dots$

2.3 Байесовский подход

Для распределения Пуассона применение формулы Байеса (6) приводит к формуле распределения плотности вероятности для параметра λ

$$P(\lambda|n) = \pi(\lambda)P(n|\lambda), \quad (92)$$

где $\pi(\lambda)$ - априорное распределение (приор) [18], удовлетворяющее тривиальному условию нормировки

$$\int_0^\infty P(\lambda|n)d\lambda = \int_0^\infty \pi(\lambda)P(n|\lambda)d\lambda = 1. \quad (93)$$

Преимуществом Байесовского подхода является то, что формула (92) по сути дела сводит задачу статистики к задаче теории вероятностей, поскольку мы знаем плотность вероятности параметра λ и можем применять методы теории вероятностей для ограничения параметра λ .

К недостаткам Байесовского подхода относятся:

1. Параметр λ не является случайным параметром в классическом частотном понятии теории вероятностей. Параметр λ имеет вполне определенное значение, просто мы его не знаем. Обычно на эту критику адепты Байесовского подхода отвечают, что формула (92) определяет степень доверия к нашему пониманию, каким может быть параметр λ .
2. Результаты расчетов в Байесовском подходе зависят от неизвестной функции $\pi(\lambda)$, которую мы должны определить вообще говоря из “ad hoc” соображений.

Часто в литературе используется либо плоский prior $\pi(\lambda) = const$, либо степенной prior $\pi(\lambda) = const \cdot \lambda^{-\beta}$, причем особенно популярен prior $\pi(\lambda) = \frac{const}{\sqrt{\lambda}}$. Заметим, что для распределения Пуассона в силу тождества $\int_0^{\infty} P(n|\lambda)d\lambda = 1$ для плоского priora $const = 1$.

Итак, в Байесовском подходе задача нахождения верхнего ограничения на параметр λ на уровне $(1 - \alpha)$ достоверности сводится к решению неравенства

$$\int_0^{\lambda} P(\lambda'|n)d\lambda' \leq 1 - \alpha \quad (94)$$

или

$$\int_{\lambda}^{\infty} P(\lambda'|n)d\lambda' \geq \alpha. \quad (95)$$

Из неравенства (95) можно получить ограничение сверху на параметр λ

$$\lambda \leq \lambda_{up}(n_{obs}, \alpha, \pi). \quad (96)$$

При этом параметр $\lambda_{up}(n_{obs}, \alpha, \pi)$ определяется из уравнения

$$\int_0^{\lambda_{up}} P(\lambda'|n)d\lambda' = 1 - \alpha. \quad (97)$$

Заметим, что неравенство (96) явно зависит от функции priora $\pi(\lambda)$. При выборе плоского priora $\pi(\lambda) = 1$ в силу тождеств

$$\int_{\lambda}^{\infty} P(n_0|\lambda')d\lambda' = \sum_{n=0}^{n_0} P(n|\lambda) \quad (98)$$

и

$$\int_0^\lambda P(n_0|\lambda')d\lambda' = \sum_{n=n_0+1}^{\infty} P(n|\lambda) \quad (99)$$

получаем, что частотное неравенство (86) совпадает с Байесовским неравенством (95). Иными словами, при выборе плоского приора $\pi(\lambda) = 1$ Байесовский подход (численные оценки) для получения верхнего предела на параметр λ в распределении Пуассона совпадает с частотным подходом, основанном на использовании формулы (86).

В случае приора $\pi(\lambda) = const \cdot \lambda^\beta$ зависимость $\lambda_{up}(n_{obs} = 0, \alpha = 0.05, \beta)$ от параметра β показана на Рис. 9.

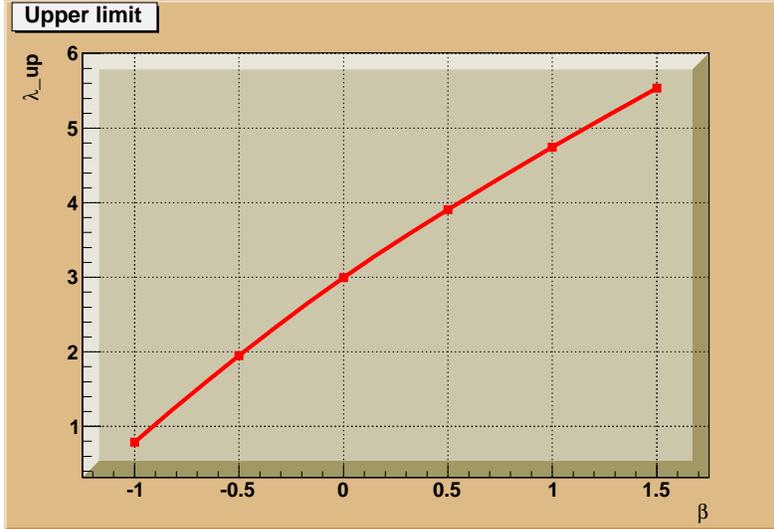


Рис. 9: Зависимость верхнего предела от степени β в выражении для степенного приора при $n_{obs} = 0$ и $\alpha = 0.05$.

Задача об определении ограничения снизу λ_{low} на параметр λ сводится к решению неравенства

$$\int_{\lambda}^{\infty} P(\lambda'|n)d\lambda' \leq 1 - \alpha. \quad (100)$$

Параметр λ_{low} определяется из уравнения

$$\int_{\lambda_{low}}^{\infty} P(\lambda'|n)d\lambda' = 1 - \alpha. \quad (101)$$

Таблица 2: Значения верхних λ_{up} и нижних λ_{low} пределов для $n_{obs} \leq 10$ и $\alpha = 0.05$, $\alpha = 0.1$. Приор $\pi(\lambda) \sim \frac{1}{\lambda}$.

n_{obs}	λ_{low}	λ_{up}	λ_{low}	λ_{up}
	$1 - \alpha = 0.9$		$1 - \alpha = 0.95$	
1	0.105	2.30	0.051	3.00
2	0.532	3.89	0.355	4.74
3	1.10	5.32	0.818	6.30
4	1.74	6.68	1.37	7.75
5	2.43	7.99	1.97	9.15
6	3.15	9.27	2.61	10.51
7	3.89	10.53	3.29	11.84
8	4.66	11.77	3.98	13.15
9	5.43	12.99	4.70	14.43
10	6.22	14.21	5.43	15.71

Заметим, что в силу тождества (98) для приора $\frac{const}{\lambda}$ результаты Байесовского и частотного подходов для λ_{low} совпадают.

Для часто используемого приора $\pi(\lambda) \sim \frac{1}{\lambda}$ [19] в Таб. 2 представлены значения верхних λ_{up} и нижних λ_{low} пределов для $n_{obs} \leq 10$ и $\alpha = 0.05$, $\alpha = 0.1$.

2.4 Метод максимального правдоподобия

В методе максимального правдоподобия [20] для случая статистики Пуассона функция максимального правдоподобия

$$L(\lambda) = P(n_{obs}|\lambda). \quad (102)$$

Из условия максимума функции правдоподобия

$$\frac{\partial}{\partial \lambda} L(\lambda) = 0 \quad (103)$$

находим наиболее вероятное значение параметра λ

$$\lambda_{max} = n_{obs}. \quad (104)$$

В методе максимального правдоподобия верхний и нижний пределы на параметр λ определяются исходя из неравенства

Таблица 3: Значения верхних λ_{up} и нижних λ_{low} пределов в методе максимального правдоподобия для $n_{obs} \leq 10$ и $\alpha = 0.05$, $\alpha = 0.1$.

n_{obs}	λ_{low}	λ_{up}	λ_{low}	λ_{up}
	$1 - \alpha = 0.9$		$1 - \alpha = 0.95$	
0	—	0.82	—	1.34
1	0.19	2.88	0.11	3.64
2	0.69	4.40	0.48	5.29
3	1.29	5.80	0.98	6.80
4	1.95	7.13	1.55	8.23
5	2.65	8.43	2.12	9.61
6	3.38	9.71	2.82	10.96
7	4.13	10.95	3.51	12.28
8	4.90	12.19	4.21	13.58
9	5.68	13.41	4.93	14.86
10	6.48	14.61	5.67	16.12

$$\sqrt{2 \ln \frac{L(\lambda_{max})}{L(\lambda)}} \leq s(\alpha). \quad (105)$$

Асимптотически условие (105) сводится к стандартному центральному интервалу с $\pm s(\alpha)$ стандартных гауссовских отклонений. Для распределения Пуассона неравенство (105) принимает вид

$$\sqrt{2[n_{obs} \ln \frac{n_{obs}}{\lambda} + \lambda - n_{obs}]} \leq s(\alpha). \quad (106)$$

Например, при $n_{obs} = 0$ верхний предел на параметр λ имеет вид

$$\lambda \leq \lambda_{up}(n_{obs} = 0, \alpha = 0.05) = 1.34. \quad (107)$$

Оценка (107) отличается от частотной оценки (88).

Как видно из Таб. 1–3 в общем случае как частотный и Байесовский подходы, так и метод максимального правдоподобия дают численно разные значения верхнего и нижнего пределов λ_{up} и λ_{low} .

3 Оценка параметров

3.1 Общая постановка задачи

Предположим у нас имеется набор экспериментальных данных $\vec{x} = (x_1, x_2, \dots, x_n)$ случайной величины X , распределенной с плотностью вероятности $f(x|\vec{\theta})$, где $\vec{\theta}$ -параметры, описывающие распределение. Задача состоит в том, чтобы оценить параметры $\vec{\theta}$ и найти интервалы доверия с доверительной вероятностью $1 - \alpha$ для полученных оценок.

В частотном подходе Неймана [16] задача решается следующим образом. С помощью функции распределения $F(x|\theta)$ мы можем найти такие функции $x_1(\theta, \alpha)$ и $x_2(\theta, \alpha)$, что справедливо равенство

$$P(x_1 < X < x_2|\theta) = 1 - \alpha = \int_{x_1}^{x_2} f(x|\theta)dx. \quad (108)$$

Смысл равенства (108) очевиден и заключается в том, что оно связывает вероятность $1 - \alpha$ нахождения случайной величины X в пределах между x_1 и x_2 с интегралом от плотности вероятности $f(x|\theta)$. Интервал $[x_1(\theta, \alpha), x_2(\theta, \alpha)]$ зависит от измеренной величины \vec{x} и объединение таких интервалов называется доверительным интервалом $D(\alpha)$. Для многих распределений (нормальное распределение, распределение Пуассона) $x_1(\theta, \alpha)$ и $x_2(\theta, \alpha)$ монотонные функции от параметра θ . Смотри в качестве иллюстрации Рис. 10.

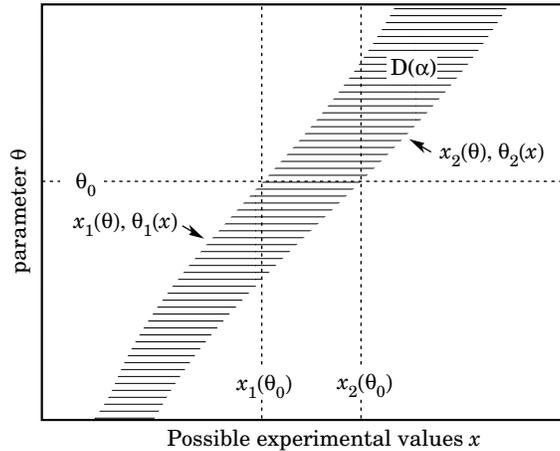


Рис. 10: Построение доверительного интервала.

В эксперименте мы измеряем случайную величину X и получаем какое-то конкретное значение x_0 .

Интервал доверия для параметра θ это набор всех значений θ для которых соответствующая линия сегмента пересекается с вертикальной линией (см. Рис. 10). Такие интервалы доверия имеют уровень достоверности $1 - \alpha$. Иными словами, справедливо равенство

$$1 - \alpha = P(x_1(\theta) < X < x_2(\theta)) = P(\theta_2(X) < \theta < \theta_1(X)), \quad (109)$$

где

$$\begin{aligned} \theta_2(X) &= \max_{\theta} x_1(\theta, x_0), \\ \theta_1(X) &= \min_{\theta} x_2(\theta, x_0). \end{aligned} \quad (110)$$

В этом вероятностном утверждении границы интервала $\theta_1(X)$ и $\theta_2(X)$ являются случайными величинами, в то время как неизвестный параметр θ есть неизвестная константа. Это означает, что если эксперимент повторять много раз, то интервалы $[\theta_1, \theta_2]$, получаемые в каждом измерении, будут различными, но доля интервалов, покрывающих истинное значение параметра θ , будет равна значению $1 - \alpha$.

Подчеркнем, что условие (108) не определяет однозначно параметры x_1, x_2 . Существует бесконечно много значений (x_1, x_2) , удовлетворяющих условию (108).

3.2 Нормальное распределение

Рассмотрим в качестве примера нормальное распределение

$$N(x|\mu, \sigma^2) \sim \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Существует много способов выбора интервалов параметров x_1 и x_2 в формуле (108).

Перечислим наиболее естественные способы.

1. $x_2(\theta) = \infty$. Этот выбор соответствует так называемому верхнему пределу.
2. $x_1(\theta) = -\infty$. Этот выбор соответствует нижнему пределу.
3. $-x_2(\theta) < X - \mu < x_2(\theta)$. Этот выбор соответствует симметричному интервалу.

Таблица 4: Соответствие для одностороннего предела между вероятностями α и значимостями s .

α	s	α	s
0.1587	1σ	0.1	1.28σ
$2.275 \cdot 10^{-2}$	2σ	0.05	1.64σ
$1.35 \cdot 10^{-3}$	3σ	0.025	1.96σ
$3.15 \cdot 10^{-5}$	4σ	0.005	2.58σ
$2.87 \cdot 10^{-7}$	5σ	0.0005	3.29σ
$1.00 \cdot 10^{-9}$	6σ	$0.5 \cdot 10^{-4}$	3.89σ

Заметим, что часто используется интервал $[x_1(\theta), x_2(\theta)]$ такой, что плотность распределения $P(x|\theta)$ больше или равна плотности распределения вне интервала $[x_1(\theta), x_2(\theta)]$ [17]⁷. В случае распределения Гаусса это определение совпадает с предыдущим симметричным определением. В общем случае, например распределение Пуассона, симметричное определение и определение, основанное на выборе интервала с максимальной плотностью, не совпадают.

Предположим, что мы знаем дисперсию σ^2 нормального измерения и хотим определить ограничения на параметр μ нормального распределения. В случае, когда мы хотим найти верхний предел на параметр μ равенство (109) записывается в виде

$$1 - \alpha = P(-\infty < X < x_2) = \int_{-\infty}^{x_2} N(x|\mu, \sigma^2) dx =$$

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x_2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x_2-\mu}{\sigma}} e^{-\frac{y^2}{2}} dy = 1 - \frac{1}{\sqrt{2\pi}} \int_{\frac{x_2-\mu}{\sigma}}^{\infty} e^{-\frac{y^2}{2}} dy. \quad (111)$$

Вводим стандартное понятие односторонней значимости как

$$\alpha = \frac{1}{\sqrt{2\pi}} \int_s^{\infty} e^{-\frac{y^2}{2}} dy. \quad (112)$$

Заметим, что обычно используемое при установлении пределов в физике высоких энергий $\alpha = 0.05$ соответствует значению $s = 1.64$. В Таб. 4 для ряда значений вероятности α приведены соответствующие им значения значимости s .

Из уравнения (111) получаем уравнение

⁷При таком выборе справедливо равенство $P(x_1|x_1(\theta)) = P(x_2|x_2(\theta))$.

$$\frac{x_2 - \mu_{up}}{\sigma} = s(\alpha) \quad (113)$$

для нахождения параметра μ , решение которого есть $\mu_{up} = x_2 + \sigma s(\alpha)$, позволяющее определить верхний предел на параметр μ

$$\mu \leq \mu_{up} = x_2 + \sigma s(\alpha) \quad (114)$$

на уровне $1 - \alpha$ достоверности.

В случае, когда мы хотим найти нижний предел на параметр μ , уравнение (111) запишется в виде

$$\begin{aligned} 1 - \alpha &= P(x_1 < X < \infty) = \int_{x_1}^{\infty} N(x|\mu, \sigma^2) dx = \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{\frac{x_1-\mu}{\sigma}}^{\infty} e^{-\frac{y^2}{2}} dy = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x_1-\mu}{\sigma}} e^{-\frac{y^2}{2}} dy. \end{aligned} \quad (115)$$

Отсюда получаем уравнение $\frac{x_1 - \mu_{low}}{\sigma} = s(\alpha)$ для определения нижнего предела $\mu_{low} = x_1 - \sigma s(\alpha)$. Таким образом, $\mu \geq \mu_{low}$.

В случае, когда мы ищем центральный интервал, который определяется из условия равенства α нижнего и верхнего пределов, то есть одновременного выполнения равенств (111) и (115), мы находим (см. Рис 11), что с вероятностью $1 - 2\alpha$

$$-\sigma s(\alpha) \leq \mu - x_1 \leq \sigma s(\alpha). \quad (116)$$

В силу симметрии нормального распределения относительно отражения $(\mu - x) \rightarrow -(\mu - x)$ верхнее и нижнее значения по модулю $(\mu - x_1)$ совпадают, то есть $|x_1 - \mu| \leq \sigma s(\alpha)$.

Заметим, что три вышеперечисленных способа определения доверительных интервалов параметра μ отнюдь не единственны. Например, мы можем требовать справедливость уравнения (108) на уровне $(1 - \alpha_1)$, уравнения (111) на уровне $(1 - \alpha_2)$ (см. Рис.12). Тогда обобщение симметричного ограничения (116) примет вид $-\sigma s(\alpha_2) \leq \mu - x_1 \leq \sigma s(\alpha_1)$.

Напомним, что в методе максимального правдоподобия функция правдоподобия зависит от набора N независимых измерений $\vec{x} = (x_1, x_2, \dots, x_N)$

$$L(\vec{\theta}) = \prod_{i=1}^N f(x_i|\vec{\theta}), \quad (117)$$

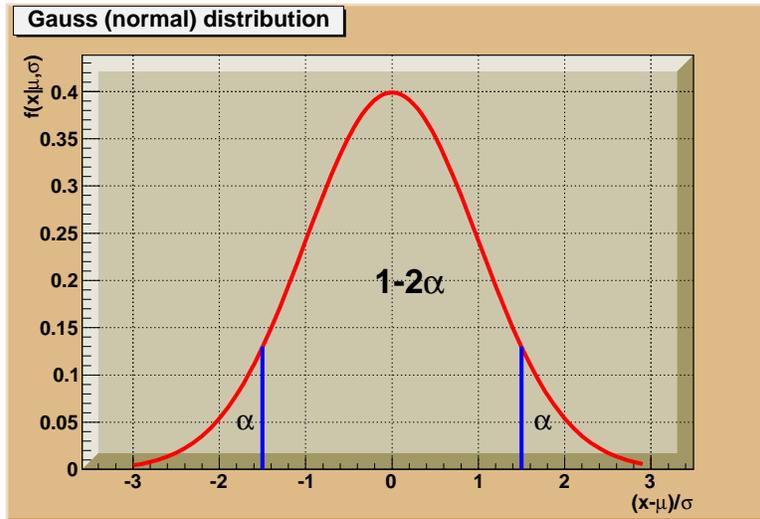


Рис. 11: На рисунке показан симметричный доверительный интервал.

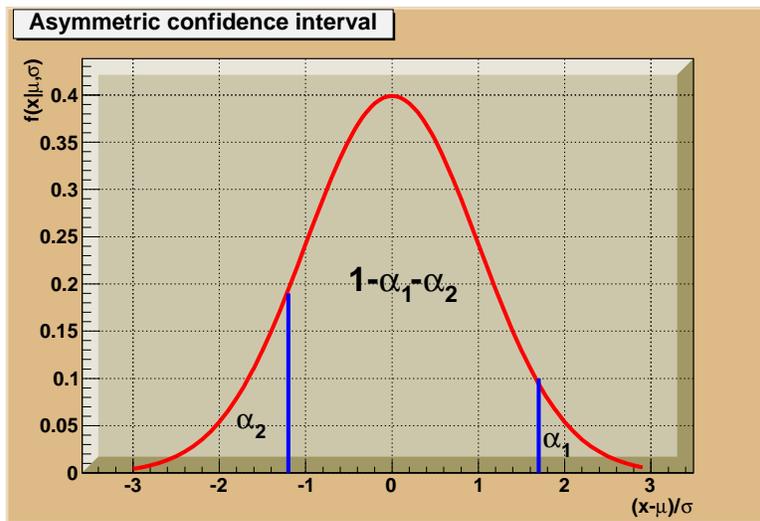


Рис. 12: На рисунке показан асимметричный доверительный интервал.

где $f(x_i|\vec{\theta})$ плотность вероятности измерения случайной наблюдаемой X .

Параметры $\vec{\theta}$ функции распределения находятся из условия максимума функции правдоподобия

$$\frac{\partial}{\partial \vec{\theta}} L(\vec{\theta}) = 0.$$

Оценки на параметры $\vec{\theta}$ находятся из условия

$$2[\ln L(\vec{\theta}_{max}) - \ln L(\vec{\theta})] \leq s^2, \quad (118)$$

где s -значимость. Фактически определение (118) выбрано так, чтобы оценка на параметры с помощью метода максимального правдоподобия и частотный подход совпадали бы для случая нормального распределения. Действительно, для случая нормального распределения с известной дисперсией σ^2 и с неизвестным средним μ уравнение (118) примет вид

$$\frac{(x_1 - \mu)^2}{\sigma^2} \leq s^2. \quad (119)$$

Решение уравнения (119)

$$-s\sigma \leq x_1 - \mu \leq s\sigma \quad (120)$$

есть ничто иное, как центральный интервал (116) в частотном подходе. В силу четности нормального распределения вероятность левого и правого хвостов совпадают.

Рассмотрим задачу определения параметра μ в Байесовском подходе. При этом мы также предполагаем, что дисперсия σ^2 известна. Плотность распределения вероятности в Байесовском подходе имеет вид

$$P(\mu|x_0, \sigma) = P(\mu)N(x_0|\mu, \sigma), \quad (121)$$

где $P(\mu)$ – неизвестный приор. При этом единственное условие на приор $P(\mu)$ это условие нормировки

$$\int P(\mu|x_0, \sigma) d\mu = 1.$$

При заданном приоре $P(\mu)$ формула (121) сводит задачу статистики к задаче теории вероятностей и мы можем определять верхние и нижние пределы на параметр μ исходя из уравнений

$$\int_{\mu_{up}}^{\infty} P(\mu|x_0, \sigma)d\mu = \alpha, \quad (122)$$

$$\int_{-\infty}^{\mu_{down}} P(\mu|x_0, \sigma)d\mu = \alpha. \quad (123)$$

Как следствие уравнений (122,123) получаем, что для центрального интервала справедливо соотношение

$$\int_{\mu_{down}}^{\mu_{up}} N(x_0|\mu, \sigma)d\mu = 1 - 2\alpha. \quad (124)$$

Заметим, что для $P(\mu) = const$

$$P(\mu|x_0, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\mu-x_0)^2}{2\sigma^2}}.$$

и, как следствие, для постоянного приора мы получаем точно такие же оценки, как и в частотном подходе.

Итак, на примере нормального распределения мы продемонстрировали три подхода к определению ограничений на параметр μ . В частотном подходе мы имеем дело с интегралами от плотности вероятности, с помощью которых получают ограничения на параметр μ . В методе максимального правдоподобия ограничения выводятся с помощью отношения плотности вероятности по отношению к максимальной плотности вероятности (формула (118)). В Байесовском подходе формула (121) сводит задачу определения ограничений на параметр μ к задаче теории вероятности, причем для приора $P(\mu) = 1$. Байесовский подход совпадает с частотным. Следует также подчеркнуть, что выбор интервала $\mu_1 \leq \mu \leq \mu_2$ неоднозначен в частотном и в Байесовском подходах, тогда как в методе максимального правдоподобия выбор интервала однозначен и он соответствует упорядочению по $\frac{\vec{\theta}}{\vec{\theta}_{max}}$ (берутся только те $\vec{\theta}$, которые дают максимум этого отношения). Вообще говоря, результаты Байесовского подхода могут сильно отличаться от частотного, но в случае постоянного приора $P(\mu) = 1$ оба подхода совпадают.

3.3 Оценка параметра в распределении Пуассона

В физике высоких энергий важную роль играет распределение Пуассона. Действительно, на БАКе количество событий с той или иной сигнатурой (например количество димюоннов, диджетов, ...) является случайной

величиной, распределенной по закону Пуассона $P(n|\lambda)$, где $\lambda = L\sigma\epsilon$ [21]. Здесь L –светимость, σ –сечение, ϵ –эффективность регистрации. Предположим, что в эксперименте мы обнаружили $n = n_{obs}$ событий. Встает вопрос, как получить ограничение на параметр λ из эксперимента?

3.3.1 Частотный подход

Для распределения Пуассона $P(n|\lambda)$ задача определения интервалов доверия $[\lambda_{low}, \lambda_{up}]$ сводится к задачам определения верхнего ограничения λ_{up} и нижнего ограничения λ_{low} . Задача получения верхнего ограничения на параметр $\lambda \leq \lambda_{up}$ на уровне $(1 - \alpha)$ достоверности сводится к решению неравенства

$$P(n \leq n_{obs}|\lambda) \equiv \sum_{n=0}^{n_{obs}} P(n|\lambda) \geq \alpha. \quad (125)$$

Из неравенства (125) получаем ограничение сверху $\lambda \leq \lambda_{up}(n_{obs}, \alpha)$ на параметр λ на уровне $(1 - \alpha)$ достоверности. В частности при $n_{obs} = 0$ $\lambda_{up} \leq ln \frac{1}{\alpha}$. Для стандартно используемой $\alpha = 0.05$ $\lambda_{up} \leq 3.0$.

Задача же получения нижнего предела на параметр λ на уровне $(1 - \beta)$ достоверности согласно частотному подходу сводится к неравенству

$$\sum_{n=n_{obs}}^{\infty} P(n|\lambda) \geq \beta. \quad (126)$$

Из неравенства (126) получаем ограничение снизу на параметр λ : $\lambda \geq \lambda_{-}(n_{obs}, \beta)$ на уровне $(1 - \beta)$ достоверности. Для $n_{obs} = 1$ неравенство (126) примет вид $\lambda_{low} \geq ln \frac{1}{1 - \beta}$ или при $\beta = 0.05$ $\lambda \geq 0.05$.

Верхние и нижние пределы λ_{low} и λ_{up} для $n_{obs} = 0, 1, 2, \dots, 10$ при $\alpha = \beta = 0.05$ и $\alpha = \beta = 0.1$ приведены в Таб. 1.

3.3.2 Метод наибольшего правдоподобия

Для распределения Пуассона функция максимального правдоподобия

$$L(\lambda) = \frac{1}{n_{obs}!} (\lambda)^{n_{obs}} e^{-\lambda}. \quad (127)$$

Решение уравнения

$$\frac{d}{d\lambda} L(\lambda) = 0 \quad (128)$$

есть $\lambda = n_{obs}$. Неравенство (118) в нашем случае запишется в виде

$$2\ln L(\lambda_{max} = n_{obs}) - 2\ln L(\lambda) \leq s^2 \quad (129)$$

или

$$2[(\lambda - n_{obs}) + n_{obs}(\ln n_{obs} - \ln \lambda)] \leq s^2. \quad (130)$$

Решение неравенства (130) представимо в виде

$$\lambda_{low}(n_{obs}, s) \leq \lambda \leq \lambda_{up}(n_{obs}, s). \quad (131)$$

Особо следует рассмотреть случай $n_{obs} = 0$. В этом случае из неравенства (131) получаем $\lambda \leq \frac{s^2}{2}$. Для $\alpha = 0.05$ следует $s = 1.64$ (см. Таблица 4) и, соответственно, $\lambda \leq \frac{s^2}{2} = 1.34$.

В Таб. (3) приведены значения λ_{low} и λ_{up} для $n_{obs} = 0, 1, 2, \dots, 10$ и для уровней достоверности $1 - \alpha = 0.95$ и $1 - \alpha = 0.9$.

Заметим, что уравнение (128) для определения интервала доверия в методе максимального правдоподобия дает тот же результат, что и частотный подход только для случая нормального распределения. В общем случае результаты могут сильно отличаться. Проиллюстрируем это на примере распределения Лапласа

$$P(x, a) = \frac{1}{2a} e^{-a|x|}, \quad \infty < x < \infty, \quad a \geq 0. \quad (132)$$

Метод максимального правдоподобия (уравнение (129)) приводит к неравенству

$$2a|x| \leq s^2(\alpha). \quad (133)$$

Тогда как частотный подход приводит к неравенству

$$e^{-a|x|} \geq \alpha \quad (134)$$

и, соответственно, $-a|x| \geq \ln \alpha$, $a|x| \leq \ln \frac{1}{\alpha}$.

Численно соотношения сильно различаются. Так, например, при $|x_0| = 1$ и $s(\alpha) = 3$ ($\alpha = 2.7 \cdot 10^{-3}$) метод максимального правдоподобия приводит к $\alpha \leq \frac{3}{2}$, тогда как частотный подход дает $\alpha \leq 5.9$.

3.3.3 Байесовский метод

Как уже отмечалось ранее в случае Байесовского подхода задача статистики сводится к задаче теории вероятностей. Для распределения Пуассона формула Байеса имеет вид

$$P(\lambda|n_{obs}) = P(\lambda)P(n_{obs}|\lambda), \quad (135)$$

где prior $P(\lambda)$ удовлетворяет условию нормировки

$$\int_0^{\infty} P(\lambda|n_{obs})d\lambda = 1, \quad (136)$$

и задача определения параметров λ_{low} и λ_{up} сводится к решению уравнения

$$\int_{\lambda_{low}}^{\lambda_{up}} P(\lambda|n_{obs})d\lambda = 1 - \alpha - \beta. \quad (137)$$

Выбор интервала $(\lambda_{low}, \lambda_{up})$ не единственный. По аналогии с частотным случаем его можно выбрать, потребовав, чтобы верхние и нижние пределы на λ определялись бы из уравнений

$$\int_{\lambda_{up}}^{\infty} P(\lambda|n_{obs})d\lambda = \alpha, \quad (138)$$

$$\int_0^{\lambda_{low}} P(\lambda|n_{obs})d\lambda = \beta. \quad (139)$$

В силу тождеств

$$\int_{\lambda_{up}}^{\infty} P(n_{obs}|\lambda)d\lambda = \sum_{n=0}^{n_{obs}} P(n|\lambda_{up}), \quad (140)$$

$$\int_0^{\lambda_{low}} P(n_{obs}|\lambda)d\lambda = \sum_{n=n_{obs}+1}^{\infty} P(n|\lambda_{low}). \quad (141)$$

верхний предел λ_{up} в Байесовском подходе совпадает с частотным верхним пределом при $P(\lambda) = const$, а нижний предел λ_{low} совпадает с частотным нижним пределом при $P(\lambda) = \frac{const}{\lambda}$.

Итак, в Байесовском подходе уравнения (137,138,139) определяют допустимую область параметра λ

$$\lambda_{low} \leq \lambda \leq \lambda_{up} \quad (142)$$

на уровне $(1 - \alpha - \beta)$ достоверности. Неравенство (142) в частотном подходе для приора $\pi(\lambda) = const$ эквивалентно неравенству

$$1 - \beta \geq \sum_{n=0}^{n_{obs}} P(n|\lambda) \geq \alpha. \quad (143)$$

Для приора $\pi(\lambda) = \frac{const}{\lambda}$ неравенство (142) эквивалентно неравенству

$$1 - \alpha \geq \sum_{n=n_{obs}}^{\infty} P(n|\lambda) \geq \beta. \quad (144)$$

Другой естественный выбор интервала $[\lambda_{low}, \lambda_{up}]$ в интеграле (137) заключается в принципе упорядочивания, а именно, точки вошедшие в интервал $[\lambda_{low}, \lambda_{up}]$ должны обладать большей плотностью распределения по сравнению с точками не вошедшими в интервал $[\lambda_{low}, \lambda_{up}]$. То есть,

$$P(\lambda \in [\lambda_{low}, \lambda_{up}]|n) \geq P(\lambda \notin [\lambda_{low}, \lambda_{up}]|n).$$

Как следствие справедливо равенство

$$P(\lambda_{low}|n) = P(\lambda_{up}|n).$$

3.4 Оценка сигнала в распределении Пуассона при ненулевом фоне

Очень часто при поиске новых явлений параметр λ а распределении Пуассона можно представить как сумму двух слагаемых

$$\lambda = \lambda_b + \lambda_s, \quad (145)$$

где $\lambda_b = L\epsilon_b\sigma_b$ представляет собой вклад от фона и предполагается в данном подразделе точно известным, а $\lambda_s = L\sigma_s\epsilon_s$ —вклад от сигнала. Задача состоит в том, чтобы оценить вклад сигнала, то есть получить ограничение на параметр λ_s . Эта задача очень похожа на рассматриваемую в предыдущем подразделе задачу об оценке параметра λ исходя из наблюдаемого значения n_{obs} сигнальных событий. Разница заключается в том, что в силу соотношений (145) и неравенств $\lambda_b \geq 0$, $\lambda_s \geq 0$ ограничение на параметр λ имеет вид $\lambda \geq \lambda_b$.

3.4.1 Частотный подход

Проблемы с определением интервалов в случае ограничения на параметры распределения. Проблемы возникают, когда в рас-

пределении $P(x, \theta)$ на параметры распределения θ налагаются дополнительные условия, следующие из физических соображений.

Рассмотрим пример [22] нормального распределения

$$N(x|\mu, \sigma = 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}, \quad (146)$$

причем параметр μ может принимать только неотрицательные значения $\mu \geq 0$ (например, μ – масса частицы). Если мы измерили $x = -1.8$, то согласно наивной процедуре извлечения параметра μ на уровне 1σ мы получаем $\mu = -1.8 \pm 1$, что находится в противоречии с требованием неотрицательности параметра μ . Принцип максимального правдоподобия с учетом ограничения $\mu \geq 0$ приводит к наиболее вероятному значению для параметра μ равному

$$\mu_{best} = \begin{cases} x_0, & x \geq 0 \\ 0, & x < 0 \end{cases} = \max(0, x_0). \quad (147)$$

Встает также вопрос о построении допустимого интервала на параметр μ . В работе [22] было предложено решение этой проблемы на основе использования упорядочивания с помощью функции максимального правдоподобия, которая в нашем случае имеет вид

$$R(\mu|x) = \frac{N(x|\mu, \sigma = 1)}{N(x|\mu_{best}, \sigma = 1)} = \begin{cases} e^{-\frac{(x-\mu)^2}{2}}, & x \geq 0 \\ e^{x\mu - \frac{\mu^2}{2}}, & x < 0 \end{cases}. \quad (148)$$

В методе максимального правдоподобия при $x_0 \leq 0$ неравенство (118) примет вид

$$(x_0 - \mu)^2 - x_0^2 \leq s^2 \quad (149)$$

или

$$0 \leq \mu \leq x_0 + \sqrt{s^2 + x_0^2}. \quad (150)$$

В случае $x_0 \geq 0$ неравенство (118) имеет стандартный вид

$$(\mu - x_0)^2 \leq s^2. \quad (151)$$

С учетом дополнительного условия $\mu \geq 0$ решение неравенства (151) есть

$$\max(0, -s + x_0) \leq \mu \leq x_0 + s. \quad (152)$$

Согласно принципу упорядочивания [22] функция $R(\mu|x)$ определяет порядок добавления точек x в допустимую область при каждом конкретном μ . А именно, добавляются точки x с максимальным значением $L(\mu|x)$. Более конкретно это означает, что для заданного значения μ находится интервал $[x_1, x_2]$ такой, что

$$R(\mu|x_1) = R(\mu|x_2),$$

$$\int_{x_1}^{x_2} N(x|\mu, \sigma = 1)dx = 1 - \alpha. \quad (153)$$

Далее используется стандартная процедура Неймана для определения допустимого интервала на значения параметров $[\mu_1, \mu_2]$ (см. Рис 13).

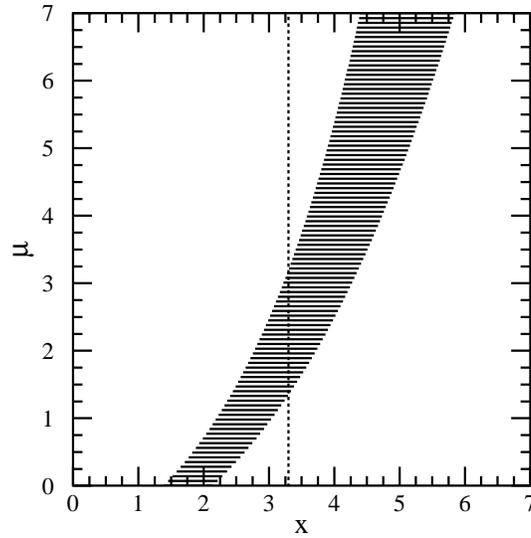


Рис. 13: Построение доверительного коридора и его использование.

Для измеренного значения x_0 параметры μ_1, μ_2 , определяющие концы интервала $[\mu_1, \mu_2]$, соответствуют уравнению (153), когда $x_1(x_2)$ берется равным $\mu_2(\mu_1)$.

В таблицах (Таб. 5,6,7 [22]) приведены конкретные значения параметров μ_1, μ_2 в зависимости от измеренного значения n_{obs} и ожидаемого фона λ_b для распределения Пуассона.

Таблица 5: Значения верхних μ_2 и нижних μ_1 пределов в методе Фельдмана-Кузинса для $n_{obs} \leq 10$ и $\alpha = 0.05$, $\alpha = 0.1$. Ожидаемый фон $\lambda_b = 0$.

n_{obs}	μ_1	μ_2	μ_1	μ_2
	$1 - \alpha = 0.9$		$1 - \alpha = 0.95$	
0	0.0	2.44	0.0	3.09
1	0.11	4.36	0.05	5.14
2	0.53	5.91	0.36	6.72
3	1.10	7.42	0.82	8.25
4	1.47	8.60	1.37	9.76
5	1.84	9.99	1.84	11.26
6	2.21	11.47	2.21	12.75
7	3.56	12.53	2.58	13.81
8	3.96	13.99	2.94	15.29
9	4.36	15.30	4.36	16.77
10	5.50	16.50	4.75	17.82

Таблица 6: Значения верхних μ_2 и нижних μ_1 пределов в методе Фельдмана-Кузинса для $n_{obs} \leq 10$ и $\alpha = 0.05$, $\alpha = 0.1$. Ожидаемый фон $\lambda_b = 1$.

n_{obs}	μ_1	μ_2	μ_1	μ_2
	$1 - \alpha = 0.9$		$1 - \alpha = 0.95$	
0	0.0	1.61	0.0	2.33
1	0.0	3.36	0.00	4.14
2	0.0	4.91	0.00	5.72
3	0.10	6.42	0.00	7.25
4	0.74	7.60	0.37	8.76
5	1.25	8.99	0.97	10.26
6	1.61	10.47	1.61	11.75
7	2.56	11.53	1.97	12.81
8	2.96	12.99	2.33	14.29
9	3.36	14.30	3.36	15.77
10	4.50	15.50	3.75	16.82

Таблица 7: Значения верхних μ_2 и нижних μ_1 пределов в методе Фельдмана-Кузинса для $n_{obs} \leq 10$ и $\alpha = 0.05$, $\alpha = 0.1$. Ожидаемый фон $\lambda_b = 3$.

n_{obs}	μ_1		μ_2	
	$1 - \alpha = 0.9$	$1 - \alpha = 0.95$	$1 - \alpha = 0.9$	$1 - \alpha = 0.95$
0	0.0	1.08	0.0	1.63
1	0.0	1.88	0.0	2.63
2	0.0	3.04	0.0	3.84
3	0.0	4.42	0.0	5.25
4	0.0	5.60	0.0	6.76
5	0.0	6.99	0.0	8.26
6	0.15	8.47	0.0	9.75
7	0.89	9.53	0.29	10.81
8	1.51	10.99	0.98	12.29
9	1.88	12.30	1.62	13.77
10	2.63	13.50	2.25	14.82

Заметим, что в рамках Байесовского подхода задача состоит в выборе приора $P(\mu)$ такого, что $P(\mu) = 0$ при $\mu < 0$. В частности, вместо приора $P(\mu) = 1$, часто используемого для нормального распределения, когда параметр μ может принимать произвольные значения, можно использовать приор $const \cdot \theta(\mu)$ автоматически учитывающий неотрицательность параметра μ ⁸. При использовании приора $p(\mu) = const \cdot \theta(\mu)$ нижний предел на параметр μ определяется с помощью соотношения

$$\frac{\int_{\mu_-}^{\infty} e^{-\frac{1}{2}(\mu-x_0)^2} d\mu}{\int_{-x_0}^{\infty} e^{-\frac{1}{2}y^2} dy} \leq 1 - \alpha. \quad (154)$$

Заметим, что в классическом частотном подходе в формулах (154) знаменатель предполагается равным 1. “Волевым образом” поделив пределы на вероятность выпадения x_0 , мы получим формулу (154)⁹.

⁸Формула (154) по сути дела есть аналог соответствующей формулы, используемой для распределения Пуассона (формула O. Helene [23]).

⁹Из условия нормировки $P(\mu|x_0) = \theta(\mu) \cdot const \cdot e^{-\frac{(\mu-x_0)^2}{2}}$, $\int P(\mu|x_0)d\mu = 1$, следует, что $const = \frac{1}{\frac{1}{\sqrt{2\pi}} \int_{-x_0}^{\infty} e^{-\frac{y^2}{2}} dy}$. Здесь $\theta(\mu) = \begin{cases} 0, & \mu = 0 \\ 1, & \mu \geq 0 \end{cases}$.

Задача определения верхнего предела на параметр μ сводится к решению неравенства

$$\frac{\int_0^{\mu+} e^{-\frac{1}{2}(\mu-x_0)^2} d\mu}{\int_{-x_0}^{\infty} e^{-\frac{1}{2}y^2} dy} \leq 1 - \alpha. \quad (155)$$

В случае распределение Пуассона $P(n|\lambda_b + \lambda_s)$ с ненулевым известным фоном λ_b функция правдоподобия максимальна при $\lambda_{s,best} = \max(0, n - \lambda_b)$.

В подходе Фельдмана-Кузинса выбор индивидуальных слагаемых в сумме

$$\sum_{n=n_-}^{n_+} P(n|\lambda_b + \lambda_s) \quad (156)$$

осуществляется согласно принципу упорядочивания по функции

$$R = \frac{P(n|\lambda_b + \lambda_s)}{P(n|\lambda_b + \lambda_{s,best})},$$

то есть допустимые члены в сумме (156) выбираются максимально большими.

Функция $R = \frac{P(n|\lambda_b + \lambda_s)}{P(n|\lambda_b + \lambda_{s,best})}$ используется для упорядочивания значений n . Далее строится интервал доверия такой, что $P(\lambda_s \in [\lambda_1, \lambda_2]) \geq 1 - \alpha$. Это осуществляется с помощью конструкции Неймана, а именно, строится сумма

$$\sum_{n=n_1}^{n_2} P(n|\lambda_b + \lambda_s) \geq 1 - \alpha. \quad (157)$$

В сумме (157) слагаемые подбираются с помощью упорядочивания по величине R . В итоге мы имеем полосу Неймана $n_1(\lambda_s)$, $n_2(\lambda_s)$ (см. Рис 13). Далее стандартным образом находятся $\lambda_{s_{low}}$ и $\lambda_{s_{up}}$ такие, что интервал доверия имеет вид $\lambda_{s_{low}} \leq \lambda_s \leq \lambda_{s_{up}}$.

Заметим, что в частотном подходе при извлечении верхнего предела на сигнал s из распределения Пуассона $P(n|\lambda_b + \lambda_s)$ и $n = n_{obs}$ на уровне $(1 - \alpha)$ достоверности числа наблюдаемых событий ограничение на λ_s получается из неравенства

$$P(n \leq n_{obs}|\lambda_b + \lambda_s) = \sum_{n=0}^{n_{obs}} P(n|\lambda_b + \lambda_s) \geq \alpha. \quad (158)$$

В модифицированном частотном подходе (CL_s подходе) уравнение (158) заменяется на

$$\frac{P(n \leq n_{obs} | \lambda_b + \lambda_s)}{P(n \leq n_{obs} | \lambda_b)} \geq \alpha, \quad (159)$$

что соответствует в Байесовском подходе замене функции приора $\theta(\lambda) \rightarrow const \cdot \theta(\lambda - \lambda_b)$.

3.4.2 Байесовский подход

В Байесовском подходе тот факт, что $\lambda \geq \lambda_b$ можно учесть самым простым способом, а именно, $P(\lambda) \rightarrow \theta(\lambda - \lambda_b)P(\lambda)$.

Введение дополнительной функции $\theta(\lambda - \lambda_b)$ изменит условия нормировки. В остальном же получение ограничений на параметр λ или λ_s абсолютно такое же, как и предыдущем подразделе. Возьмем постоянный приор, для которого, как было показано, верхний Байесовский предел совпадает с частотным верхним пределом благодаря равенству (98). С учетом модификации $P(\lambda) \rightarrow \theta(\lambda - \lambda_b)P(\lambda)$ уравнение для определения верхнего предела для приора $\pi(\lambda) = const \cdot \theta(\lambda - \lambda_b)$ запишется в виде

$$\frac{\int_{\lambda_{up}}^{\infty} P(n_{obs} | \lambda) d\lambda}{\int_{\lambda_b}^{\infty} P(n_{obs} | \lambda) d\lambda} \geq \alpha \quad (160)$$

или

$$\frac{\int_{\lambda_{up} - \lambda_b = \lambda_s}^{\infty} P(n_{obs} | \lambda_b + \lambda'_s) d\lambda'_s}{\int_{\lambda_b}^{\infty} P(n_{obs} | \lambda) d\lambda} \geq \alpha. \quad (161)$$

В силу тождеств (98,99) ограничение (161) можно переписать в виде

$$\frac{\sum_{n=0}^{n_{obs}} P(n_{obs} | \lambda_b + \lambda_s)}{\sum_{n=0}^{n_{obs}} P(n_{obs} | \lambda_b)} \geq \alpha. \quad (162)$$

Для $n_{obs} = 0$ неравенство примет вид

$$\frac{e^{-\lambda_b - \lambda_s}}{e^{-\lambda_b}} \geq \alpha,$$

то есть $\lambda_s \leq \ln \frac{1}{\alpha}$. Без знаменателя отличного от единицы в формуле (162) мы имели бы неравенство $\lambda_b + \lambda_s \leq \ln \frac{1}{\alpha}$.

Аналогичное неравенство можно получить и для нижнего предела. Для определения нижнего предела для приора $\frac{const}{\lambda}$ неравенство для определения нижнего предела примет вид

$$\frac{\int_{\lambda_b}^{\lambda_b + \lambda_s} P(n_{obs} - 1 | \lambda) d\lambda}{\int_{\lambda_b}^{\infty} P(n_{obs} - 1 | \lambda) d\lambda} \geq \beta. \quad (163)$$

Для $n_{obs} = 1$ получаем $\frac{e^{-\lambda_b} - e^{-\lambda_b - \lambda_s}}{e^{-\lambda_b}} = \beta$, $1 - e^{-\lambda_s} = \beta$, $\lambda_s = \ln \frac{1}{1-\beta} \approx \beta$.

При $\beta = 0.05$ ограничение имеет вид $\lambda_s > 0.05$.

3.4.3 Оценка параметров в методе наименьших квадратов

Часто используют метод наименьших квадратов, основанный на минимизации квадратичной формы. Рассмотрим выборку $\vec{y} = (y_1, y_2, \dots, y_n)$ наблюдений при измерении случайной величины Y . Пусть математическое ожидание в каждом измерении есть $E(Y_i, \vec{\theta})$, а матрица вторых моментов соответствующих распределений есть $[V]$. Здесь $\vec{\theta}$ неизвестные параметры, а $V_{ij}(\vec{\theta})$ известные функции от θ . В методе наименьших квадратов оценками параметров $\vec{\theta}$ служат такие значения $\vec{\hat{\theta}}$, которые обращают в минимум квадратичную форму

$$Q^2 = \sum_{i=1}^N \sum_{j=1}^N [y_i - E(Y_i, \vec{\theta})] ([V]^{-1})_{ij} [y_j - E(Y_j, \vec{\theta})] = [\vec{y} - E(\vec{y}, \vec{\theta})]^T [V]^{-1} [\vec{y} - E(\vec{y}, \vec{\theta})]. \quad (164)$$

В случае отсутствия корреляций ковариантная матрица $V_{ii} = \sigma_i^2(\vec{\theta})$ диагональна и выражение (164) к известной сумме квадратов

$$Q^2 = \sum_{i=1}^N \sigma_i^{-2}(\vec{\theta}) [Y_i - E(Y_i, \vec{\theta})]^2.$$

В случае нормального N -мерного распределения метод максимального правдоподобия и метод наименьших квадратов совпадают.

Метод наименьших квадратов часто используется при фитировании гистограмм. А именно, при выборе бинов в которых достаточно большое количество событий, так что приближение нормального распределения оправдано и в приближении отсутствует корреляция между бинами, необходимо минимизировать величину

$$Q^2 = \sum_{i=1}^N \frac{(N_i - \lambda_i(\vec{\theta}))^2}{\lambda_i(\vec{\theta})}. \quad (165)$$

Это есть не что иное как обычный χ^2 -метод, применяемый для фита гистограмм.

Для практических приложений вместо минимизации величины (165) более удобно минимизировать величину

$$Q_{mod}^2 = \sum_{i=1}^N \frac{(N_i - \lambda_i(\vec{\theta}))^2}{N_i}. \quad (166)$$

Поскольку $E(y_i) = \lambda_i(\vec{\theta})$, модифицированное выражение (166) является разумным приближением при минимизации (164).

Можно так же в методе максимального правдоподобия максимизировать непосредственно

$$\ln L = \sum_{i=1}^N N_i \ln \lambda_i(\vec{\theta}) - \sum_{i=1}^N \lambda_i(\vec{\theta}).$$

В приближении $\ln L \approx \sum_{i=1}^N N_i \ln \lambda_i(\vec{\theta})$ ($\sum \lambda_i(\vec{\theta}) = const$) задача сводится

к более простой задаче максимизации $\sum_{i=1}^N N_i \ln \lambda_i(\vec{\theta})$.

3.4.4 Интервал доверия в случае нескольких параметров

Пусть существует оценка $\hat{\vec{\theta}}$ параметров $\vec{\theta}$ и эта оценка распределена по нормальному многомерному закону со средним $\vec{\theta}$ и матрицей вторых моментов $[V]$. В этом случае

$$f(\hat{\vec{\theta}}|\vec{\theta}) = \frac{1}{(2\pi)^{\frac{N}{2}} (\det[V])^{\frac{1}{2}}} e^{-\frac{1}{2}(\hat{\vec{\theta}}-\vec{\theta})^T [V]^{-1} (\hat{\vec{\theta}}-\vec{\theta})}. \quad (167)$$

Ковариантная форма $Q(\hat{\vec{\theta}}, \vec{\theta}) = (\hat{\vec{\theta}}-\vec{\theta})^T [V]^{-1} (\hat{\vec{\theta}}-\vec{\theta})$ имеет $\chi^2(N)$ -распределение (см. раздел 1.7.6). Это означает в частности, что распределение $Q(\hat{\vec{\theta}}, \vec{\theta})$ не зависит от $\vec{\theta}$ и можно написать вероятностное утверждение

$$P(Q(\hat{\vec{\theta}}, \vec{\theta}) \leq K_\beta^2) = \beta, \quad (168)$$

где β -вероятность того, что оценка параметра будет находиться в области доверия, определяемой точкой K_β^2 для $\chi^2(N)$ -распределения.

Область в пространстве $\vec{\theta}$, задаваемая уравнением $Q(\hat{\vec{\theta}}, \vec{\theta}) = K_\beta^2$, имеет вид гиперэллипсоида, поверхность которого соответствует постоянно значению плотности вероятности для функции (167).

В двумерном случае матрица $[V]$ это

$$[V] = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Эллипс содержит параметры, одновременно удовлетворяющие неравенству (168).

Оценка методом максимального правдоподобия в случае нескольких параметров Обобщение одномерного неравенства (129) имеет вид $2[\ln L(\vec{\theta}_{max}) - \ln L(\vec{\theta})] \leq \chi_{\beta}^2(k)$.

Как показано (например, [20]), функция $-2\ln l(\vec{\theta})$, где $l(\vec{\theta}) = \frac{L(\vec{\theta})}{L(\vec{\theta}_{max})}$, распределена асимптотически как $\chi^2(k)$. Это приводит к неравенству $\beta = P[-2\ln l(\vec{\theta}) \leq \chi_{\beta}^2(k)]$, которое определяет интервал доверия для параметров $\vec{\theta}$.

3.4.5 Асимптотическая нормальность оценки максимального правдоподобия

Оценка параметра распределения $\hat{\theta}$ по методу максимального правдоподобия асимптотически распределена как $N(\theta, \frac{1}{NI})$, где I – информация о параметре θ в одном измерении, а N – число измерений [6]. Отсюда следует, что интервал доверия для $\hat{\theta}$ имеет асимптотический вид $\hat{\theta} \pm \frac{\lambda_{\frac{\beta}{2}}}{\sqrt{NI(\theta)}}$, где $\lambda_{\frac{\beta}{2}}$ есть $\frac{\beta}{2}$ точка стандартного нормального распределения.

Более точная оценка получается из того факта, что $\frac{\partial}{\partial\theta} \ln L$ распределено со средним значением равным нулю и вариацией $NI(\theta)$. Распределение $\frac{1}{\sqrt{NI(\theta)}} \frac{\partial}{\partial\theta} \ln L$ является асимптотически нормальным распределением. Доверительный интервал определяется неравенством $|\frac{1}{\sqrt{NI(\theta)}} \frac{\partial}{\partial\theta} \ln L| \leq \lambda_{\frac{\beta}{2}}$.

3.5 Ожидаемые пределы

В условиях будущего эксперимента мы не знаем количество событий n в распределении Пуассона $P(n|\lambda)$. Как правило мы знаем (с какой-то точностью) параметр $\lambda = \lambda_b$ для СМ. В случае присутствия новых взаимодействий

$$\lambda = \lambda_b + \lambda_s,$$

где λ_s – вклад физики вне рамок СМ. Поэтому возникает естественный вопрос – какое ограничение сверху на параметр λ_s мы можем получить в будущем эксперименте? Вопрос сводится к следующему: какое значение n_{obs} в соответствующих формулах (Глава 3) необходимо использовать? Наиболее распространенный рецепт (“Azimov dataset”) это использование наиболее вероятного значения $n_{obs} = [\lambda_b]$ для распределения Пуассона $P(n|\lambda)$. Так, например, ограничение (162)¹⁰ примет вид

$$\frac{\sum_{n=0}^{[\lambda_b]} P([\lambda_b]|\lambda_b + \lambda_s)}{\sum_{n=0}^{[\lambda_b]} P([\lambda_b]|\lambda_b)} \geq \alpha.$$

Смысл анзаца $n_{obs} = [\lambda_b]$ очень простой. Вероятности того, что $n_{obs} \leq [\lambda_b]$ или $n_{obs} > [\lambda_b]$ приблизительно равны 50%. И, поэтому, с вероятностью 50% в случае реального эксперимента можно получить как лучшее, так и худшее ограничение, чем ожидаемое ограничение.

В случае измерения непрерывной величины, распределенной по нормальному закону $N(x|\mu_b, \sigma^2)$, наиболее вероятное значение случайной переменной $x = \mu_b$. Ограничение на параметр сигнала μ_s в предположении, что случайная переменная X распределена в этом случае по нормальному закону $N(x|\mu_b + \mu_s, \sigma^2)$ на уровне 1σ есть $-\sigma \leq \mu_s \leq \sigma$.

В случае нормального распределения $N(x|\mu_b, \sigma^2)$ в 68% случаев (1σ) случайная величина X должна лежать в интервале $\mu_b - \sigma \leq X \leq \mu_b + \sigma$. Отсюда мы получаем, что в 68% случаев ограничение на параметр новой физики μ_s на уровне 1σ лежит в интервале $-2 \sigma \leq \mu_s \leq 2 \sigma$.

Для случая распределения Пуассона обобщение вышеприведенной процедуры очевидно и состоит в выборе таких значений n , лежащих вблизи $[\lambda_b]$, что

$$\sum_{n_k} P(n_k|\lambda_b) \geq 68\%.$$

Далее для каждого n_k находим верхний предел на параметр сигнала λ_s ($\lambda = \lambda_s + \lambda_b$) и для каждого λ_b выбираются наибольшие (n_k^{max}) и наименьшие (n_k^{min}) значения n_k и с их помощью определяются соответственно наименее “удачное” и наиболее “удачное” значения верхнего предела на λ_s .

¹⁰Напомним, что ограничение (162) естественно получается в Байесовском подходе и часто именуется CL_S методом [24].

4 Учет систематических ошибок

4.1 Введение

Типичная задача в физике высоких энергий это извлечение из экспериментальных данных информации о возможности существования новой физики вне рамок СМ. Например, при изучении той или иной реакции¹¹ с наложением ряда кинематических ограничений (обрезаний) вероятность детектирования n событий определяется распределением Пуассона $P(n|\lambda)$, где $\lambda = \lambda_b + \lambda_s$, $\lambda_b = L\epsilon_b\sigma_b$, $\lambda_s = L\epsilon_s\sigma_s$. Здесь L – полная интегральная светимость, ϵ_b , ϵ_s – эффективность детектирования для, соответственно, фоновых и сигнальных событий, σ_b , σ_s сечения фона и сигнала.

Сечение сигнала σ_s зависит от параметров модели, описывающих новое явление (массы суперсимметричных частиц, масса Z' -бозона, ...). Основная задача состоит в нахождении из экспериментальных данных величины или ограничения на сигнальное сечение σ_s или (что почти то же самое) в нахождении параметров (или ограничений на них), определяющих новую физику.

Проблема заключается в том, что параметры модели $L, \epsilon_b, \epsilon_s, \sigma_b, \sigma_s$ определяются с какой-то точностью, это называется систематическими эффектами. Так, например, точность определения интегральной светимости L в процессе работы БАКа в 2010 году оценивается равной 11% [25].

Параметры эффективности регистрации фона и сигнала зависят от работы детектора и связаны главным образом с не 100% эффективностью регистрации частиц, а также с неточным определением импульса частиц.

При определении σ_b и σ_s возникают неопределенности, связанные с неточным теоретическим знанием сечений. Здесь, правда следует сказать, что иногда величину фона σ_b можно извлечь из экспериментальных данных в другой кинематической области.

Обычно систематические неопределенности разделяют на три класса:

1. систематические неопределенности, которые можно устранить путем измерения величин в другой кинематической области. При этом систематические неопределенности такого рода определяют

¹¹ Например, количество димюонных событий с ограничениями по поперечным импульсам мюонов и их инвариантной массе.

ся статистическими неопределенностями измеряемых величин. Например, при определении сечения фона сигнатуры с $n \geq 1 jets + E_T^{miss}$ с адронными струями и ненулевым поперечным потеряннм импульсом, фон, возникающий от реакции $n \geq 1 jets + (Z \rightarrow \nu\bar{\nu})$, можно оценить исходя из измерения сечения процесса $n \geq 1 jets + (Z \rightarrow l^+l^-)$ и известной величины $\frac{\Gamma(Z \rightarrow \nu\bar{\nu})}{\Gamma(Z \rightarrow l^+l^-)}$.

2. Неопределенности, связанные с модельными предположениями при извлечении данных. Предположения об эффективности регистрации частиц, точности определения их импульсов, мисидентификации (инструментальные фоны) ¹².
3. Неопределенности, связанные с неточным определением теоретических сечений σ_b и σ_s . Как уже отмечалось выше в ряде случаев, но не всегда, определить сечение фона σ_b можно исходя из экспериментальных данных в другой кинематической области.

Существует несколько способов учета систематических эффектов. Они будут рассмотрены в последующих трех подразделах.

4.2 Оценка параметра фона, исходя из измерений в другой кинематической области

Проблему можно сформулировать следующим образом: пусть интересующая нас для поиска новой физики кинематическая область описывается распределением Пуассона $P(n_1, \lambda_b + \lambda_s)$. Связанная с ней кинематическая область, где как мы предполагаем сигнал очень мал, описывается распределением Пуассона $P(n_2, \tau\lambda_b)$, где τ – известное число. Задача состоит в том, чтобы по измеренному значению n_{2obs} определить параметр λ_b и с помощью измерения числа событий n_{1obs} в интересующей нас кинематической области получить ограничение на параметр сигнала λ_s .

Решение этой задачи наиболее просто в методе максимального правдоподобия. Функция правдоподобия равна

$$L(\lambda_b, \lambda_s | n_{1obs}, n_{2obs}) = P(n_{2obs} | \lambda_b + \lambda_s) P(n_{1obs} | \tau\lambda_b). \quad (169)$$

Далее из условия максимума

$$\frac{\partial}{\partial \lambda_b} L(\lambda_b, \lambda_s) = \frac{\partial}{\partial \lambda_s} L(\lambda_b, \lambda_s) = 0 \quad (170)$$

¹²Типичный пример мисидентификации – это неправильное отождествление адронной струи в качестве электрона или, что намного реже, в качестве мюона.

находятся наиболее правдоподобные значения параметров λ_b и λ_s и получаются ограничения на λ_b и λ_s .

Заметим, что в Байесовском подходе, как уже отмечалось выше, справедлива формула

$$P(\lambda_s | n_{2obs}) = \int d\lambda_b P(n_{2obs} | \lambda_b + \lambda_s) \pi(\lambda_b). \quad (171)$$

А приор $\pi(\lambda_b)$ с учетом измерения n_{2obs} для фона можно определить как

$$\pi(\lambda_b) = \frac{P(n_{1obs} | \tau \lambda_b) \eta(\lambda_b)}{\int d\lambda_b P(n_{1obs}, \tau \lambda_b) \eta(\lambda_b)}, \quad (172)$$

где $\eta(\lambda_b)$ – приор, используемый при измерении фонового процесса.

4.3 Оценка систематических неопределенностей в методе максимального правдоподобия

Предположим, что мы измеряем количество событий, распределенных по закону Пуассона $P(n | \lambda_b + \lambda_s)$, и неопределенность фона λ_b описывается нормальным распределением

$$P(\lambda_b | \lambda_b^0, \Delta \lambda_b^0) = \frac{1}{\sqrt{2\pi} \Delta \lambda_b^0} e^{-\frac{1}{2} \left(\frac{\lambda_b - \lambda_b^0}{\Delta \lambda_b^0} \right)^2}. \quad (173)$$

Функция правдоподобия имеет вид

$$L(\lambda_s, \lambda_b | \lambda_b^0, n) = \frac{(\lambda_b + \lambda_s)^n}{n!} e^{-\lambda_b - \lambda_s} \frac{e^{-\frac{1}{2} \left(\frac{\lambda_b - \lambda_b^0}{\Delta \lambda_b^0} \right)^2}}{\sqrt{2\pi} \Delta \lambda_b^0}. \quad (174)$$

Оценка параметров для случая отсутствия сигнала ($\lambda_s = 0$) получается путем решения уравнения

$$\frac{\partial}{\partial \lambda_b} L(\lambda_s = 0, \lambda_b | \lambda_b^0, n) = 0. \quad (175)$$

Решение имеет вид

$$\lambda_b(\lambda_b^0, \Delta \lambda_b^0, n) = \frac{\lambda_b^0 - (\Delta \lambda_b^0)^2}{2} + \sqrt{\left(\frac{\lambda_b^0 - (\Delta \lambda_b^0)^2}{2} \right)^2 + n (\Delta \lambda_b^0)^2}. \quad (176)$$

Значение вероятности P осуществления n_{obs} измерений в отсутствие сигнала определяется стандартной формулой

$$P(\lambda_b^0 | n_{obs}) = P[N \geq n_{obs} | \lambda_b = \lambda_b(\lambda_b^0, \Delta\lambda_b^0, n)] = \sum_{k=n_{obs}}^{\infty} \frac{(\lambda_b(\lambda_b^0, \Delta\lambda_b^0, n))^k e^{-\lambda_b(\lambda_b^0, \Delta\lambda_b^0, n)}}{k!}. \quad (177)$$

В случае ненулевого λ_s из условия максимума функции правдоподобия $L(\cdot)$ по отношению к переменным λ_s и λ_b

$$\frac{\partial}{\partial \lambda_s} L(\cdot) = \frac{\partial}{\partial \lambda_b} L(\cdot) = 0 \quad (178)$$

находится функция максимального правдоподобия $L(\hat{\lambda}_s, \hat{\lambda}_b | \lambda_b^0, \lambda_s^0, n)$ и сравнивается с функцией максимального правдоподобия в отсутствии сигнала. Далее, используя методы изложенные в Главе (3), можно получить ограничение на параметр сигнала λ_s .

4.4 Метод усреднения Кузинса-Хайланда

Метод Кузинса-Хайланда по существу применяет Байесовский подход к проблеме систематики или несущественных параметров.

Предположим, что нам известен Байесовский приор $\pi(\theta')$ для всех несущественных параметров θ' (например, параметров, описывающих неопределенности сечения фона и сигнала). Для функции распределения $P(x|\theta, \theta')$, где x -случайная величина, θ -существенные параметры, проинтегрируем по несущественным параметрам θ' и получим усредненную функцию распределения

$$P_{av}(x|\theta) = \int d\theta' \pi(\theta') P(x|\theta, \theta') d\theta'. \quad (179)$$

Далее, при извлечении из экспериментальных данных информации о существенных величинах θ , используем усредненную функцию распределения $P_{av}(x|\theta)$. При извлечении из экспериментальных данных на основе усредненной функции распределения $P_{av}(x|\theta)$ можно применять как частотный, так и Байесовский подходы, а также метод максимального правдоподобия.

Как уже отмечалось ранее, параметры λ_b и λ_s представимы в виде $\lambda_b = L\epsilon_b\sigma_b$, $\lambda_s = L\epsilon_s\sigma_s$. При оценке систематической неопределенности, например, λ_b параметра необходимо знать распределения параметров $L, \epsilon_b, \sigma_b, \sigma_s$.

Обычно предполагается, что параметры $L, \epsilon_b, \epsilon_s, \sigma_b, \sigma_s$ независимы и подчиняются нормальному распределению¹³.

¹³Часто при оценках используют также логнормальное распределение.

Проиллюстрируем метод Кузинса-Хайланда на примере измерения случайной величины X , распределенной по нормальному закону $N(x|\mu, \sigma^2)$. Пусть мы точно знаем дисперсию, а величина μ распределена по нормальному закону $N(\mu|\mu_0, \sigma_\mu^2)$. Здесь μ_0 – теоретическое (среднее) значение величины μ , а σ_μ^2 – дисперсия, отражающая неточное знание параметра μ . Наша задача состоит в получении ограничений на параметр μ_0 . Напомним, что согласно методу Кузинса-Хайланда распределение случайной величины X определяется путем интегрирования по несущественным параметрам

$$P_{av}(x|\theta) = \int d\theta' \pi(\theta') P(x|\theta, \theta') d\theta' \quad (180)$$

В нашем случае формула (180) принимает вид

$$G(x|\mu_0, \sigma^2) = \int_{-\infty}^{\infty} d\mu N(x|\mu, \sigma^2) N(\mu|\mu_0, \sigma_\mu^2) = N(x|\mu_0, \sigma^2 + \sigma_\mu^2). \quad (181)$$

Таким образом, учет систематики, связанной с неточным знанием величины μ , сводится к замене дисперсии

$$\sigma^2 \rightarrow \sigma^2 + \sigma_\mu^2. \quad (182)$$

Предположим, что мы измерили величину X равную x_0 . Тогда, с учетом замены (182), значимость нашего измерения определяется стандартной формулой

$$s = \frac{|x_0 - \mu_0|}{\sqrt{\sigma^2 + \sigma_\mu^2}}. \quad (183)$$

Отличие систематической ошибки σ_μ^2 , связанной с неточным знанием величины μ_0 , от статистической ошибки заключается в том, что она, как правило, пропорциональна среднему значению μ_0 , то есть $\sigma_\mu^2 \sim \mu_0^2$.

В случае N измерений случайной величины X среднее значение $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$ с учетом систематики распределена по нормальному закону

$N(\bar{x}|\mu_0, \sigma_{\bar{x}}^2)$, причем $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{N} + \sigma_\mu^2$.

Дисперсия σ_μ^2 не зависит от числа измерений и, как следствие, получаем, что в пределе бесконечного числа измерений $\sigma_{\bar{x}}^2 \rightarrow \sigma_\mu^2$,

то есть все равно остается неопределенность в нахождении параметра μ_0 . И так, в пределе большого числа измерений систематическая неопределенность доминирует и не зависит от числа измерений.

Для задачи извлечения сигнального сечения по экспериментальным данным на основе использования распределения Пуассона $P(n|\lambda)$ в случае $n_{obs} \gg 1$ распределение Пуассона аппроксимируется нормальным распределением $N(n|\lambda, \lambda)$. Для модели с $\lambda = \lambda_b + \lambda_s$, $\lambda_b = L\epsilon_b\sigma_b$, $\lambda_s = L\epsilon_s\sigma_s$ независимые неопределенности L , ϵ_b , ϵ_s , σ_b , σ_s удобно представить в виде $L = L_0(1 \pm \delta_L)$, $\epsilon_b = \epsilon_b^0(1 \pm \delta_{\epsilon_b})$, $\epsilon_s = \epsilon_s^0(1 \pm \delta_{\epsilon_s})$, $\sigma_b = \sigma_b^0(1 \pm \delta_{\sigma_b})$, $\sigma_s = \sigma_s^0(1 \pm \delta_{\sigma_s})$.

В случае малых неопределенностей δ_L , δ_{ϵ_b} , δ_{ϵ_s} уравнение для определения сигнального сечения σ_s^0 можно записать в виде

$$\sigma_s^0 = \frac{n_{obs} - L_0\epsilon_b^0\sigma_b^0}{L_0\epsilon_s^0} \pm x, \quad (184)$$

причем ошибка измерения сигнального сечения σ_s^0 равна

$$x = \sqrt{\left(\frac{\epsilon_b^0\sigma_b^0}{\epsilon_s^0}\right)^2(\delta_{\epsilon_s}^2 + \delta_{\epsilon_b}^2 + \delta_{\sigma_b}^2) + \frac{n_{obs}}{(L_0\epsilon_s^0)^2}(1 + n_{obs}(\delta_L^2 + \delta_{\epsilon_s}^2))}. \quad (185)$$

Таким образом, в приближении нормального распределения комбинированный учет различных систематических ошибок сводится к квадратурованию ошибок и тривиален. Проблема немного усложняется, когда количество n_{obs} невелико и мы не можем использовать приближение нормального распределения.

Пусть, например, у нас имеется распределение Пуассона $P(n|\lambda_b + \lambda_s)$, а параметры λ_b и λ_s распределены по нормальному закону

$$N(\lambda_b|\lambda_b^0, \sigma_b) = \frac{1}{\sqrt{2\pi}\sigma_b} e^{-\frac{(\lambda_b - \lambda_b^0)^2}{2\sigma_b^2}}, \quad (186)$$

$$N(\lambda_s|\lambda_s^0, \sigma_s) = \frac{1}{\sqrt{2\pi}\sigma_s} e^{-\frac{(\lambda_s - \lambda_s^0)^2}{2\sigma_s^2}}. \quad (187)$$

Согласно методу Кузинса-Хайланда мы усредняем распределение Пуассона с функциями $N(\lambda_b|\lambda_b^0, \sigma_b)$ и $N(\lambda_s|\lambda_s^0, \sigma_s)$

$$P_{av}(n|\lambda_b^0, \lambda_s^0) = \int d\lambda_b d\lambda_s P(n|\lambda_b + \lambda_s) N(\lambda_b|\lambda_b^0, \sigma_b) N(\lambda_s|\lambda_s^0, \sigma_s). \quad (188)$$

В пределе малых дисперсий σ_b и σ_s

$$\lim_{\substack{\sigma_b \rightarrow 0 \\ \sigma_s \rightarrow 0}} P_{av}(n|\lambda_b^0, \lambda_s^0, \sigma_b, \sigma_s) = P(n|\lambda_b^0, \lambda_s^0). \quad (189)$$

Разумеется, очень важный вопрос, как распределения или приоры $P(\lambda_b|\lambda_b^0, \sigma_b)$, $P(\lambda_s|\lambda_s^0, \sigma_s)$ выбирать для учета систематики. При малых дисперсиях σ_b , σ_s это несущественно. При больших дисперсиях результат усреднения начинает зависеть от формы распределения. Кроме того, из физических соображений $\lambda_b \geq 0$ и $\lambda_s \geq 0$, в то время как нормальное распределение приводит к возможным значениям λ_b и λ_s от $-\infty$ до $+\infty$. При небольших дисперсиях эта проблема решается волевым путем – обнулением плотностей вероятностей при отрицательных λ_b и λ_s и изменением нормировки с тем, чтобы полный интеграл по λ_b , λ_s от 0 до ∞ равнялся бы 1. Часто используется логнормальное по λ_b , λ_s распределение, для которого переменные $\ln \lambda_b$, $\ln \lambda_s$ автоматически лежат от $-\infty$ до $+\infty$.

В работе [26] изучалось влияние систематики на извлечение верхнего ограничения на параметр λ распределения Пуассона $P(n, \lambda)$. В предположении, что систематическая неопределенность в параметре λ распределена по нормальному закону с дисперсией σ_λ , при малых σ_λ учет систематики приводит к ограничению $\lambda \leq \lambda_{up}$, где $\lambda_{up} = \lambda_{up}^0(1 + \frac{(\lambda_{up}^0 - n_{obs})\sigma_\lambda^2}{2})$ и λ_{up}^0 – верхний предел без учета систематики. При $n_{obs} = 0$ и 90% уровне достоверности $\lambda_{up}(n_{obs} = 0, \sigma_\lambda) = 2.30(1 + 2.30\frac{\sigma_\lambda^2}{2})$.

Как уже отмечалось выше параметр λ в распределении Пуассона $P(n, \lambda)$ представим в виде $\lambda = L\epsilon\sigma$ и мы должны учитывать систематические ошибки, связанные с неточным знанием светимости L , эффективности регистрации ϵ и сечения σ . Предположим, что ошибки распределены по нормальному закону и они маленькие. Мы можем символически записать $L = L_0(1 \pm \delta_L)$, $\epsilon = \epsilon_0(1 \pm \delta_\epsilon)$, $\sigma = \sigma_0(1 \pm \delta_\sigma)$, где δ_L , δ_ϵ , $\delta_\sigma \ll 1$ – стандартные отклонения. Предполагая, что ошибки δ_L , δ_ϵ , δ_σ независимы, получаем, что неопределенность параметра λ распределена по нормальному закону с дисперсией $\delta_\lambda = \sqrt{\delta_L^2 + \delta_\epsilon^2 + \delta_\sigma^2}$, то есть $\lambda = \lambda_0(1 \pm \delta_\lambda)$, где $\lambda_0 = L_0\epsilon_0\sigma_0$.

Таблица 8: Верхние пределы λ_{up} при 90% уровне доверия для значений $n_{obs} = 0, 1, 2, 3$ и значений $\delta_\lambda = 0, 0.1, 0.2$.

$\delta_\lambda \setminus n_{obs}$	0	1	2	3
0.0	2.30	3.89	5.32	6.68
0.1	2.33	3.95	5.41	6.80
0.2	2.41	4.11	5.67	7.17

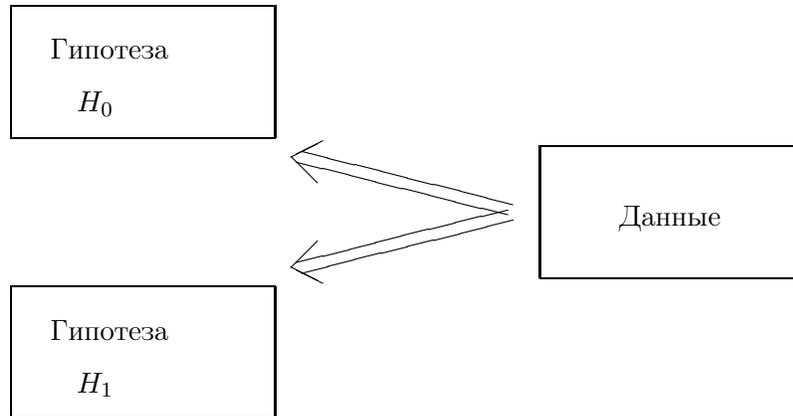
В таблице (8) показано влияние систематики на значение верхнего предела на параметр распределения λ с помощью частотного метода [26] при неопределенности параметра λ в 0, 10 и 20% при $n_{obs} \leq 3$.

5 Проверка гипотез

5.1 Введение

Одной из основных задач любого эксперимента является проверка той или иной модели (теории), предназначенной для поиска экспериментальных подтверждений, а также поиск новых явлений. В настоящее время вся совокупность экспериментальных данных в физике высоких энергий описывается Стандартной Моделью [27]¹⁴. Стандартная Модель (СМ) предсказывает существование бозона Хиггса – последней неоткрытой частицы СМ. Существуют многочисленные расширения СМ (суперсимметрия, дополнительные размерности, Z' -бозоны, ...), предсказывающие существование новой физики. Основной задачей БАКа является поиск бозона Хиггса и новой физики вне рамок СМ. Это делается путем детектирования частиц в конечном состоянии (адронных струй, электронов, мюонов, τ -лептонов) с определенными ограничениями на импульсы частиц в конечном состоянии. В этом случае у нас есть две гипотезы.

1. Гипотеза H_0 – все описывается СМ.
2. Гипотеза H_1 – существует новая физика вне рамок СМ.



Гипотеза H_0 называется простой гипотезой, если она не имеет свободных параметров. Это хорошая идеализация для СМ без учета вкладов от распадов бозона Хиггса, поскольку точное значение массы бозона Хиггса не известно. Следует подчеркнуть, что в реальной жизни существуют неопределенности, связанные, в частности, с неуменьением точно

¹⁴Единственный экспериментальный факт, лежащий вне предсказаний Стандартной Модели – осцилляции нейтрино [28]. Заметим, что нейтринные осцилляции легко объясняются путем включения масс нейтрино в Стандартную Модель.

предсказывать сечения СМ, а также экспериментальные неопределенности, связанные с неточным знанием полной светимости, неточным знанием импульсов частиц и эффективности их регистрации. Поэтому понятие простой гипотезы является всего лишь разумной идеализацией. Предположим, что в качестве альтернативы гипотезе H_0 мы имеем простую гипотезу H_1 . Например, гипотезу о существовании суперсимметрии [29] с вполне определенными массами суперчастиц. Заметим, что гипотеза о существовании новой физики является как правило составной, то есть зависящей от неопределенных параметров (масс суперчастиц, параметров дополнительных размерностей, масс Z' -бозона и т.д.). Поэтому у нас на первом этапе стоит задача проверки основной гипотезы H_0 – справедливости СМ.

5.2 Проверка основной гипотезы

Предположим мы проверяем СМ в процессах рождения лептонных пар на БАКе, накладывая определенные ограничения на импульсы лептонов и их инвариантные массы. СМ при заданной светимости предсказывает вполне определенное среднее количество событий λ_b .

Вероятность наблюдения n событий описывается распределением Пуассона $P(n|\lambda)$. Предположим, что мы обнаружили n_{obs} событий и наша задача – определить насколько наблюдаемое количество событий n_{obs} совместимо с нашей основной гипотезой, что $\lambda = \lambda_b$ в распределении Пуассона.

Вероятность наблюдения $n \geq n_{obs}$ событий равна ¹⁵

$$P_+(n_{obs}|\lambda_b) \equiv P(n \geq n_{obs}|\lambda_b) = \sum_{n=n_{obs}}^{\infty} P(n|\lambda_b). \quad (190)$$

Вероятность наблюдения $n \leq n_{obs}$ событий равна

$$P_-(n_{obs}|\lambda_b) \equiv P(n \leq n_{obs}|\lambda_b) = \sum_{n=0}^{n=n_{obs}} P(n|\lambda_b). \quad (191)$$

Напомним, что $P(n_{obs}|\lambda_b)$ максимальна [30] при $n_{obs} = [\lambda]$ ¹⁶. Поэтому,

¹⁵Вследствие формулы (99) справедливо равенство

$$P_+(n_{obs}|\lambda_b) = \int_0^{\lambda_b} P(n_{obs} - 1|\lambda') d\lambda',$$

позволяющее связать частотное определение p -вероятности с Байесовским определением вероятности параметра $\lambda \leq \lambda_b$ для приора $\pi(\lambda) = \frac{const}{\lambda}$.

¹⁶Здесь квадратные скобки $[..]$ означают целую часть от числа.

Таблица 9: Значения значимостей S и соответствующие им значения вероятностей P_+ .

S	0	1	2	3	4	5
P_+	0.5	0.1587	$2.275 \cdot 10^{-2}$	$1.35 \cdot 10^{-3}$	$3.15 \cdot 10^{-5}$	$2.87 \cdot 10^{-7}$

Таблица 10: Значения λ_b для наблюдаемого числа n_{obs} , соответствующие уровню открытия $S = 5$.

n_{obs}	1	2	3	4	5	6	7	8	9	10
λ_b	0.00000028	0.00075	0.012	0.05	0.13	0.25	0.41	0.61	0.84	1.11

если наблюдаемое количество событий n_{obs} сильно отличается от $[\lambda_b]$ в ту или иную сторону, иными словами либо P_+ , либо P_- очень малы, то можно говорить о том, что СМ с ее предсказанием параметра λ_b распределения Пуассона не верна, то есть гипотеза H_0 противоречит экспериментальным данным. Поскольку в поисках новой физики, в основном, предсказывается избыток событий по сравнению с предсказаниями СМ, то есть $n_{obs} > \lambda_b$, то мы рассмотрим ситуацию, когда P_+ мало. Обычно значение вероятности P связывается со стандартной значимостью S согласно формуле

$$P = \frac{1}{\sqrt{2\pi}} \int_S^\infty e^{-\frac{x^2}{2}} dx = \frac{1}{2} [1 - \text{erf}(\frac{S}{\sqrt{2}})]. \quad (192)$$

Причем в физике высоких энергий общепринято считать, что при значимости $S \geq 5$ мы можем объявлять об открытии нового явления (новой физики вне рамок СМ). Значимость $S = 5$ соответствует вероятности $P = 2.9 \cdot 10^{-7}$ того, что явление описывается СМ¹⁷. В Таблице 9 приведены несколько значений значимостей S и соответствующих им вероятностей P_+ (см. также Таблицу 4).

В Таблице 10 приведены значения λ_b в зависимости от n_{obs} , соответствующие уровню открытия $S = 5$.

Заметим, что в реальной жизни мы знаем параметр СМ λ_b с какой-то точностью, что связано с нашим неумением точно вычислить параметр λ_b , точно измерять импульсы частиц и их идентифицировать. Как было показано в предыдущей главе, неопределенности можно учесть с помощью функции распределения $P(\lambda_b | \lambda_b^0, \sigma)$, которая обычно выбирается в

¹⁷Обычно $S \geq 3$ считается указанием в пользу новой физики.

виде нормального распределения

$$P(\lambda_b|\lambda_b^0, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\lambda_b-\lambda_b^0)^2}{2\sigma^2}}. \quad (193)$$

Учет неопределенности, связанной с неточным значением λ_b производится путем усреднения в правой части формул (190,191) с $P(\lambda_b|\lambda_b^0, \sigma)$, а именно, обобщение соответствующих формул имеет вид

$$P'_+(n_{obs}|\lambda_b^0, \sigma) = \int P_+(n_{obs}|\lambda_b, \sigma)P(\lambda_b|\lambda_b^0, \sigma)d\lambda_b, \quad (194)$$

$$P'_-(n_{obs}|\lambda_b^0, \sigma) = \int P_-(n_{obs}|\lambda_b, \sigma)P(\lambda_b|\lambda_b^0, \sigma)d\lambda_b. \quad (195)$$

Заметим, что в пределе $\sigma \rightarrow 0$ $P'_+ \rightarrow P_+$, $P'_- \rightarrow P_-$. При измерении случайной величины X , распределенной по нормальному закону

$$P(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (196)$$

вероятность

$$P(x > \mu + S\sigma) = \int_{\mu+S\sigma}^{\infty} P(x|\mu, \sigma)dx = \frac{1}{\sqrt{2\pi}} \int_S^{\infty} e^{-\frac{x^2}{2}} dx \quad (197)$$

и $S = 5$ как раз соответствует порогу открытия.

В общем случае, когда мы собираем данные X для проверки гипотезы H_0 , соответствующей распределению $f(x|\theta)$ задача состоит в нахождении тестовой статистики $T(x)$ такой, что большие значения $t_{obs} \equiv T(x_{obs})$ являются свидетельством (доказательством) против гипотезы H_0 . С этой целью вводится вероятность наблюдения $T \geq t_{obs}$

$$P = P(T \geq t_{obs}|H_0). \quad (198)$$

Как уже отмечалось выше в физике высоких энергий гипотеза H_0 отвергается, если $P \leq 2.9 \cdot 10^{-7}$.

В методе максимального правдоподобия аналог значимости S определяется формулой

$$S = \sqrt{2 \ln \frac{f(x_{obs}|\theta_{max})}{f(x_{obs}|\theta)}}. \quad (199)$$

В случае нормального распределения $N(x|\mu, \sigma^2)$ формула (199) совпадает со стандартной формулой

Таблица 11: Значения значимостей S для распределения Пуассона, вычисленные по формулам (190,192) для различных значений n_{obs} и λ_b . В скобках приведено значение значимости, вычисленное по формуле (200).

n_{obs}	$\lambda_b = 1$	$\lambda_b = 2$	$\lambda_b = 3$	$\lambda_b = 4$	$\lambda_b = 5$
1	-0.34 (0)	-1.10 (0.71)	-1.65 (1.16)	-2.09 (1.5)	-2.47 (1.79)
2	0.63 (1)	-0.24 (0)	-0.85 (0.58)	-1.33 (1.0)	-1.75 (1.34)
3	1.40 (2)	0.46 (0.71)	-0.19 (0)	-0.71 (0.5)	-1.15 (0.89)
4	2.08 (3)	1.07 (1.41)	0.38 (0.58)	-0.17 (0)	-0.63 (0.45)
5	2.68 (4)	1.62 (2.21)	0.90 (1.16)	0.33 (0.5)	-0.15 (0)
6	3.24 (5)	2.13 (2.83)	1.38 (1.73)	0.79 (1.0)	0.30 (0.45)
7	3.77 (6)	2.61 (3.54)	1.83 (2.31)	1.22 (1.5)	0.71 (0.89)
8	4.26 (7)	3.06 (4.24)	2.26 (2.89)	1.63 (2.0)	1.11 (1.34)
9	4.73 (8)	3.91 (4.95)	3.06 (3.46)	2.03 (2.5)	1.49 (1.79)
10	5.18 (9)	4.31 (5.66)	3.44 (4.04)	2.40 (3.0)	1.86 (2.23)

$$S = \frac{|x_{obs} - \mu|}{\sigma}. \quad (200)$$

В Таблице 11 приведены значения S для нескольких значений λ_b и n_{obs} . Для распределения Пуассона при малых n_{obs} различие между значением значимости (192) и значимости (200) может быть весьма существенным (Таблица 11).

В Байесовском подходе вероятность того, что параметр λ распределения Пуассона меньше или равен λ_b определяется формулой

$$P = \int_0^{\lambda_b} P(\lambda|n_{obs})d\lambda.$$

Малость параметра P означает, что основная гипотеза не верна.

5.3 Проверка сложной гипотезы

Предположим мы хотим решить, какая из гипотез H_0 (СМ физика) или H_1 (физика вне рамок СМ) правильная на основе измерения величины X . Пусть $T(x)$ некоторая функция от наблюдений, которую будем называть тестовой статистикой и пусть W будет пространством всех значений $T(x)$. Пространство W обычно разделяют на критическую область ω , когда гипотеза H_0 отвергается, и область $W - \omega$, в которой гипотеза H_0 не отвергается.

Вероятность попадания в область ω , когда гипотеза H_0 верна

$$P(T \in \omega | H_0) = \alpha \quad (201)$$

называется ошибкой первого рода. Иными словами α это вероятность того, что гипотеза H_0 отвергается, когда она верна.

Важно, чтобы выбранная нами тестовая статистика T позволяла бы различить гипотезы H_0 и H_1 . Степень способности статистики T различить гипотезы H_0 и H_1 определяется мощностью теста, определяемого как вероятность выполнения гипотезы H_1 при условии, что данные лежат в критической области

$$P(T \in \omega | H_1) = 1 - \beta. \quad (202)$$

Другими словами β это вероятность того, что T входит в допустимую область $W - \omega$ для гипотезы H_0

$$P(T \in W - \omega | H_1) = \beta. \quad (203)$$

Итак, при тестировании двух гипотез вводятся два типа ошибок:

1. Ошибка первого рода α – вероятность отвергнуть основную (нулевую) гипотезу H_0 , когда она верна.
2. Ошибка второго рода β – вероятность принятия основной (нулевой) гипотезы H_0 , когда она не верна.

Предположим гипотезы H_0 и H_1 характеризуются параметрами θ_0 и θ_1 и, соответственно, функцией распределения $f(x, \theta)$.

Ранее была определена мощность теста $P(\theta_1)$, позволяющая различать гипотезы $\theta = \theta_0$ и $\theta = \theta_1$, а именно, $P(\theta_1) = 1 - \beta_1$.

В общем случае составной гипотезы (то есть гипотезы, зависящей от параметра) мощность теста определяют как $P(\theta) = 1 - \beta(\theta)$ для альтернативной гипотезы, определяемой параметром θ .

Различные тесты можно сравнивать на основе сравнения функции мощности теста $P(\theta)$.

В случае, если альтернативная гипотеза H_1 -простая (все параметры распределения заданы) тест является максимальным, если $P(\theta_1)$ максимален. Тест, который является наиболее мощным при всех значениях θ называется однородно наиболее мощным.

Тест называется состоятельным, если при наборе данных вероятность ошибки второго рода стремится к нулю, то есть

$$\lim_{N \rightarrow \infty} P(T(x) \in \omega | H_1) = 1$$

Для распределения Пуассона с гипотезой $H_0 : \lambda = \lambda_b$ и альтернативной гипотезой $H_1 : \lambda = \lambda_b + \lambda_s$ в частотном подходе параметры α и $1 - \beta$ равны соответственно

$$\alpha = \sum_{n=n_{obs}}^{\infty} P(n|\lambda_b),$$

$$1 - \beta = \sum_{n=n_{obs}}^{\infty} P(n|\lambda_b + \lambda_s).$$

Предположим, что параметры λ_b и λ_s известны точно. Тогда в случае $\alpha \leq 2.9 \cdot 10^{-7}$ как это общепринято в физике высоких энергий – гипотеза H_0 отвергается (СМ физика не в состоянии описать экспериментальные данные). В случае, когда β меньше определенного значения (например, $\beta \leq 0.95$) можно сказать, что на 95% уровне достоверности гипотеза H_1 не противоречит экспериментальным данным.

5.4 Тест Неймана-Пирсона

Проблему нахождения наиболее мощного теста гипотезы H_0 относительно гипотезы H_1 можно сформулировать как нахождение наилучшей критической подобласти в пространстве x возможных значений наблюдений. Из определений (201,202,203) следует, что

$$\int_{\omega_\alpha} f(\vec{x}|\theta_0) d\vec{x} = \alpha, \quad (204)$$

$$\int_{\omega_\alpha} f(\vec{x}|\theta_1) d\vec{x} = 1 - \beta. \quad (205)$$

Уравнения (204,205) можно представить в виде

$$1 - \beta = E_{\omega_\alpha} \left(\frac{f(\vec{x}|\theta_1)}{f(\vec{x}|\theta_0)} \Big|_{\theta=\theta_0} \right). \quad (206)$$

Отсюда следует, что $1 - \beta$ будет максимально, если ω_α содержит долю x -данных таких x , что отношение $\frac{f(\vec{x}|\theta_1)}{f(\vec{x}|\theta_0)}$ максимально. Поэтому наилучшая критическая область ω_α состоит из точек, удовлетворяющих неравенству

$$\frac{f(\vec{x}, \theta_1)}{f(\vec{x}, \theta_0)} > c_\alpha, \quad (207)$$

причем c_α выбирается таким образом, чтобы соотношение (204) выполнялось.

Итак, если выполняется соотношение (207), выбираем гипотезу H_1 , в противном случае выбираем гипотезу H_0 . Эта процедура известна как тест Неймана-Пирсона. При этом тестовая статистика это отношение функций правдоподобия гипотез H_1 и H_0 .

В случае проверки двух сложных гипотез H_0 и H_1 , функции распределения которых зависят соответственно от параметров $\vec{\theta}_0$ и $\vec{\theta}_1$, можно определить отношение максимального правдоподобия как

$$L = \frac{\max_{\vec{\theta}_0} L(\vec{x}, \vec{\theta}_0)}{\max_{\vec{\theta}_1} L(\vec{x}, \vec{\theta}_1)} \quad (208)$$

и это отношение позволяет разделять две сложные гипотезы.

Для случая нормального распределения $N(x|\mu, \sigma^2)$, когда гипотеза H_0 определяется как $\mu = \mu_b$, $\sigma = \sigma_0$, а гипотеза H_1 соответствует $\mu = \mu_b + \mu_s$, $\sigma = \sigma_0$, отношение функций правдоподобия для измеренного значения $x = x_{obs}$ есть

$$\frac{N(x_{obs}|\mu_b + \mu_s, \sigma_0)}{N(x_{obs}|\mu_b, \sigma_0)} = e^{-\frac{1}{2\sigma_0^2}[(x_{obs} - \mu_b - \mu_s)^2 - (x_{obs} - \mu_b)^2]}.$$

Подчеркнем, что в физике высоких энергий гипотеза H_1 как правило составная (параметр λ зависит от неизвестных масс частиц и параметров взаимодействий) и нам важно продемонстрировать на первом этапе исследования, что гипотеза H_1 обладает допустимыми значениями параметров, при которых она не противоречит экспериментальным данным, что, вообще говоря, нетривиально.

Так, например, если $\mu_s \leq \sigma$, а $x_{obs} = \mu_b + 5\sigma_0$, то основная гипотеза $\mu_s = 0$ исключена на уровне 5σ ($\alpha = 2.9 \cdot 10^{-7}$), а альтернативная составная гипотеза исключена (или имеет согласие с экспериментом на уровне 4σ ($\alpha = 3.2 \cdot 10^{-5}$)), что также позволяет по сути дела исключить альтернативную гипотезу H_1 . В случае, когда $\mu_s \leq 1\sigma$, альтернативная гипотеза H_1 имеет согласие с экспериментом на уровне 1σ ($\alpha = 0.16$), что позволяет рассматривать ее как альтернативную замену основной гипотезе H_0 .

5.5 Байесовский подход к проверке гипотез

Рассмотрим для начала проверку простой гипотезы H_0 на примере распределения Пуассона $P(n|\lambda)$.

В силу формулы Байеса

$$P(\lambda|n_{obs}) = \pi(\lambda)P(n_{obs}|\lambda), \quad (209)$$

вероятность того, что $\lambda \geq \lambda_b$

$$P(\lambda \geq \lambda_b) = \int_{\lambda_b}^{\infty} P(\lambda|n_{obs})d\lambda. \quad (210)$$

При плоском приоре $\pi(\lambda) = 1$ в силу соотношений (97,98) Байесовское определение совпадает с частотным определением (87). Таким образом, в Байесовском подходе в случае известного приора $\pi(\lambda)$ нам необходимо вычислить интеграл (210). Нулевая гипотеза H_0 (параметр λ в распределении Пуассона равен λ_b) отвергается, если $P(\lambda \geq \lambda_b) \leq \alpha$. Здесь мы неявно предполагаем, что реализуется ситуация с наблюдением избытка событий. В общем случае нулевая гипотеза $H_0 : \lambda = \lambda_b$ отвергается, если $P(\lambda \geq \lambda_b) \leq \alpha$, либо $P(\lambda \leq \lambda_b) \leq \alpha$.

Напомним, что в физике высоких энергий общепринято полагать $\alpha = 2.9 \cdot 10^{-7}$.

В случае, когда мы должны сделать выбор между простыми гипотезами H_0 и H_1 , теорема Байеса приводит к следующим формулам для вероятности:

$$\pi(H_0|x) = \frac{P(x|H_0)\pi_0}{P(x|H_0)\pi_0 + P(x|H_1)\pi_1}, \quad (211)$$

$$\pi(H_1|x) = 1 - \pi(H_0|x), \quad (212)$$

где π_i – априорные вероятности для гипотез H_i , $i = 1, 2$, $P(x|H_0)$, $P(x|H_1)$ – плотности вероятности события X при гипотезах H_0 и H_1 соответственно.

Если $\pi(H_0|x) < \pi(H_1|x)$, то гипотеза H_1 более предпочтительна. В случае $\pi(H_0|x) > \pi(H_1|x)$ гипотеза H_0 более предпочтительна. Если же $\pi(H_0|x) = \pi(H_1|x)$, то невозможно сделать выбор между гипотезами H_0 и H_1 . В случае $\pi_0 = \pi_1 = \frac{1}{2}$ отношение

$$\frac{\pi(H_0|x)}{\pi(H_1|x)} = \frac{P(x|H_0)}{P(x|H_1)}. \quad (213)$$

Отношение (213) есть не что иное как отношение функций правдоподобия.

Часто гипотезы H_0 и H_1 зависят от неизвестных параметров θ (маргинальных параметров систематики). В этом случае обобщение формул (211,212) имеют вид

$$\pi(H_0|x) = \frac{\int P(x|\theta, H_0)\pi(\theta|H_0)\pi_0 d\theta}{\int [P(x|\theta, H_0)\pi(\theta|H_0)\pi_0 + P(x|\theta, H_1)\pi(\theta|H_1)\pi_1] d\theta}, \quad (214)$$

$$\pi(H_1|x) = 1 - \pi(H_0|x). \quad (215)$$

Для $\pi_0 = \pi_1 = \frac{1}{2}$ отношение вероятностей равно

$$\frac{\pi(H_0|x)}{\pi(H_1|x)} = \frac{\int P(x|\theta, H_0)\pi(\theta|H_0)\pi_0 d\theta}{\int P(x|\theta, H_1)\pi(\theta|H_1)\pi_1 d\theta}. \quad (216)$$

6 Комбинирование результатов

В случае измерения нескольких независимых величин стоит задача о комбинировании этих величин, то есть извлечения дополнительной информации по сравнению с единичными измерениями. Существует несколько способов это сделать.

6.1 Комбинирование двух нормальных распределений

В случае измерения двух независимых случайных величин X_1 и X_2 , распределенных по нормальному закону $X_1 \sim N(\mu_1, \sigma)$ и $X_2 \sim N(\mu_2, \sigma)$ с одной и той же дисперсией σ и неизвестными параметрами μ_1 и μ_2 полезно перейти к новым координатам

$$Y_1 = \frac{X_1\mu_1 + X_2\mu_2}{\sqrt{\mu_1^2 + \mu_2^2}}, \quad (217)$$

$$Y_2 = \frac{-X_1\mu_2 + X_2\mu_1}{\sqrt{\mu_1^2 + \mu_2^2}}. \quad (218)$$

Нетрудно показать, что для $X_1 \cdot X_2 \sim N(\mu_1, \sigma) \cdot N(\mu_2, \sigma)$ и $Y_1 \cdot Y_2 \sim N(\sqrt{\mu_1^2 + \mu_2^2}, \sigma) \cdot N(0, \sigma)$ справедливо

$$N(\mu_1, \sigma) \cdot N(\mu_2, \sigma) = N(\sqrt{\mu_1^2 + \mu_2^2}, \sigma) \cdot N(0, \sigma). \quad (219)$$

Для модели с $\mu_1 = \mu_{b1} + \mu_s$, $\mu_2 = \mu_{b2} + \tau\mu_s$, где τ , μ_{b1} и μ_{b2} – заданные величины, а μ_s неизвестно, необходимо комбинировать $x'_1 = x_1 - \mu_{b1}$, $x'_2 = x_2 - \mu_{b2}$. Тогда

$$Y'_1 = \frac{X'_1 + \tau X'_2}{\sqrt{1 + \tau^2}}, \quad (220)$$

$$Y'_2 = \frac{-X'_1\tau + X'_2}{\sqrt{1 + \tau^2}} \quad (221)$$

и для $X'_1 \cdot X'_2 \sim N(\mu_1 - \mu_{b1}, \sigma) \cdot N(\mu_2 - \mu_{b2}, \sigma)$ и $Y'_1 \cdot Y'_2 \sim N(\mu_s\sqrt{1 + \tau^2}, \sigma) \cdot N(0, \sigma)$.

Отсюда мы находим наиболее вероятное значение

$$\mu_s^{max} = \frac{y_1}{\sqrt{1 + \tau^2}} = \frac{(x_1 - \mu_{b1}) + \tau(x_2 - \mu_{b2})}{\sqrt{1 + \tau^2}}. \quad (222)$$

Дисперсия для параметра μ_s равна $\frac{\sigma}{\sqrt{1+\tau^2}}$. Иными словами, на 67% уровне достоверности $|\mu_s - \mu_s^{max}| \leq \frac{\sigma}{\sqrt{1+\tau^2}}$.

Заодно по распределению величины $Y_2 \sim N(0, \sigma)$ мы можем проверить справедливость гипотезы о том, что $\mu_1 = \mu_{b1} + \mu_s$ и $\mu_2 = \mu_{b2} + \tau\mu_s$.

Заметим, что принцип максимального правдоподобия $\frac{\partial}{\partial \mu_s} \ln L = 0$, где

$$L = P(x_1 | \mu_{b1} + \mu_s, \sigma) \cdot P(x_2 | \mu_{b2} + \tau\mu_s, \sigma), \quad P(x | \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

приводит к тому же значению μ_s (222), что и описанная выше процедура.

6.2 Нормальное распределение. Общий случай

Рассмотрим задачу комбинирования n независимых величин X_k , $k = 1, 2, \dots, n$, распределенных по нормальному закону $X_k \sim N(\mu_k, \sigma_k)$.

Для случая, когда все дисперсии равны $\sigma_k = \sigma_l = \sigma$, $k \neq l$ и $\mu_k = \mu_{bk} + s \cdot c_k$, где μ_{bk} – вклад фона, а s – сигнал, причем $\sum c_k^2 = 1$, нетрудно показать, что случайная величина

$$Y_1 = \sum_{k=1}^N c_k (x_k - \mu_{bk})$$

распределена по нормальному закону $Y_1 \sim N(s, \sigma)$. Ортогональные к Y_1 комбинации Y_2, Y_3, \dots, Y_n распределены по нормальному закону $Y_k \sim N(0, \sigma)$ ($k = 2, 3, \dots, n$) со средними значениями $\mu_k = 0$. Отсюда мы можем извлечь комбинированную оценку на параметр s на уровне 1σ , а именно:

$$s = s_0 \pm \sigma, \\ s_0 = \sum_{k=1}^N c_k (x_k^0 - \mu_{bk}),$$

где x_k^0 – измеренные значения случайной величины X_k^0 . Заметим, что использование метода максимального правдоподобия приводит к такому же результату. К такому же результату приводит и Байесовские вычисления с использованием функции правдоподобия

$$\prod_{k=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_k - \mu_k)^2}{2\sigma^2}}$$

и плоского приора $\pi(\lambda) = const$.

6.3 Метод наименьших квадратов

Предположим мы измеряем N независимых величин X_i , распределенных по нормальному закону с дисперсиями σ_i^2 и средними $x_{i,av}$. Пусть имеется модель, предсказывающая величины средних $x_{i,av}$ как функции от неизвестных параметров модели θ_k , ($k = 1, 2, \dots, l$). Для определения степени достоверности нашей модели согласно методу наименьших квадратов введем величину

$$T(\vec{\theta}) = \sum_{i=1}^N \frac{(x_{i,meas} - x_{i,av}(\vec{\theta}))^2}{\sigma_i^2}, \quad (223)$$

где $x_{i,meas}$ -измеренные значения случайных величин X_i . Параметры $\vec{\theta} = (\theta_1, \theta_2, \dots, \theta_l)$ определяются исходя из минимума величины $T(\vec{\theta})$ ¹⁸

$$\frac{\partial}{\partial \theta_i} T(\vec{\theta}) = 0, \quad i = 1, 2, \dots, l. \quad (224)$$

Эта конструкция очень полезна при анализе данных, поскольку величина $T(\vec{\theta})$ в случае совпадения $x_{i,av}(\vec{\theta}^{min})$ с действительными средними $x_{i,meas}$ распределена как $\chi^2(T, N)$ с N степенями свободы.

Напомним, что при больших N величина $Z_N = \frac{\chi^2(N) - N}{\sqrt{2N}}$ распределена по нормальному закону с дисперсией $\sigma = 1$ и нулевым средним.

В более общем случае в методе наименьших квадратов нам необходимо минимизировать билинейную форму

$$\chi^2 = \sum_{i,j=1}^N (m_i - M_i(\vec{\theta}))_i \sigma_{ij}^{-1} (m_j - M_j(\vec{\theta}))_j = (\vec{m} - \vec{M}(\vec{\theta}))^T C^{-1} (\vec{m} - \vec{M}(\vec{\theta})). \quad (225)$$

Здесь m_i -результаты измерений, а $M_i(\vec{\theta})$ -теоретические предсказания, зависящие от l неизвестных параметров $\vec{\theta} = (\theta_1, \dots, \theta_l)$, по которым и происходит минимизация билинейной формы.

Например, в случае $N = 2$ и ненулевого корреляционного коэффициента $\rho \neq 0$ нам необходимо минимизировать форму

$$\chi^2 = (m - m_1, m - m_2) \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}^{-1} \begin{pmatrix} m - m_1 \\ m - m_2 \end{pmatrix}. \quad (226)$$

¹⁸Здесь мы предполагаем, что дисперсии σ_k не зависят от неизвестных параметров θ_k , $k = 1, \dots, l$.

Минимизация формы (226) приводит к

$$m = \frac{m_1(\sigma_2^2 - \rho\sigma_1\sigma_2) + m_2(\sigma_1^2 - \rho\sigma_1\sigma_2)}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}, \quad (227)$$

$$\sigma_m^2 = \frac{\sigma_1^2\sigma_2^2(1 - \rho^2)}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}. \quad (228)$$

6.4 Метод максимального правдоподобия

Метод наименьших квадратов естественно получается в методе максимального правдоподобия. В случае N независимых измерений функция правдоподобия есть произведение функций правдоподобия

$$L_{tot}(\vec{\theta}) = \prod_{k=1}^N L(\vec{\theta}), \quad L(\vec{\theta}) = F_k(x_k|\vec{\theta}). \quad (229)$$

Параметры $\vec{\theta} = (\theta_1, \dots, \theta_l)$ определяются исходя из условия максимизации $L_{tot}(\vec{\theta})$, то есть

$$\frac{\partial}{\partial \theta_i} L_{tot}(\vec{\theta}) = 0, \quad i = 1, \dots, l. \quad (230)$$

В случае нормального распределения мы получаем, как следствие, условие минимума функции χ^2 (формула 225). Таким образом видно, что обобщением метода наименьших квадратов является метод максимального правдоподобия.

В случае измерения N независимых величин n_k , распределенных согласно закону Пуассона с параметрами $\lambda_k(\vec{\theta})$, зависящими от неизвестных величин $\vec{\theta}$ модели, функция правдоподобия

$$L_{tot}(\vec{\theta}) = \prod_{k=1}^N L_k(\vec{\theta}), \quad L_k(\vec{\theta}) = P(n_k^{obs}|\lambda_k(\vec{\theta})) \quad (231)$$

и неизвестные параметры $\vec{\theta}$ определяются стандартным образом как максимумы функции $L_{tot}(\vec{\theta})$.

Предположим, что мы имеем гистограмму, разбитую на N бинов, распределение событий в каждом из бинов описывается распределением Пуассона с параметрами $\lambda_i(\vec{\theta})$, где $\vec{\theta}$ -параметры теоретической модели. Тогда в методе максимального правдоподобия

$$-2 \ln L = -2 \ln \left(\prod_{i=1}^N e^{-\lambda_i(\vec{\theta})} \frac{\lambda_i(\vec{\theta})^{n_i}}{n_i!} \right). \quad (232)$$

В этом случае χ^2 -функция есть

$$\chi^2 = -2 \ln \frac{L(\lambda_i)}{L(\lambda_i^{max})} = 2 \sum_{i=1}^N [\lambda_i(\vec{\theta}) - n_i + n_i \ln \frac{n_i}{\lambda_i(\vec{\theta})}]. \quad (233)$$

Функцию χ^2 -можно представить в виде (в случае аппроксимации распределения Пуассона нормальным распределением)

$$\chi^2 = \sum_i \frac{(n_i - f_i(\vec{\theta}))^2}{n_i} - 2 \ln L - n \ln 2n - \sum_i \ln n_i. \quad (234)$$

В случае замены $n_i \rightarrow \mu_i = f_i(\vec{\theta})$ мы получаем χ^2 -распределение Пирсона

$$\chi^2 = \sum \frac{(n_i - f_i(\vec{\theta}))^2}{f_i(\vec{\theta})} \quad (235)$$

для минимизации.

С помощью χ^2 -распределения можно комбинировать данные. Например, пусть у нас есть два измерения с некоррелированными ошибками: $\mu_1 \pm \sigma_1$ и $\mu_2 \pm \sigma_2$. Тогда χ^2 -функция равна

$$\chi^2 = \frac{(\mu - \mu_1)^2}{\sigma_1^2} + \frac{(\mu - \mu_2)^2}{\sigma_2^2} \quad (236)$$

и минимизация $\frac{\partial}{\partial \mu} \chi^2 = 0$ дает оценку

$$\mu = \frac{\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}. \quad (237)$$

При этом ошибка равна

$$\frac{1}{\sigma_m^2} = -\frac{\partial^2}{\partial \mu^2} \ln L = \frac{1}{2} \frac{\partial^2}{\partial \mu^2} \chi^2 = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}. \quad (238)$$

6.5 Комбинирование пределов

Как уже было отмечено выше комбинирование результатов (скажем двух экспериментов ATLAS и CMS) достигается с помощью объединенной функции правдоподобия, которая есть произведение функций правдоподобия экспериментов ATLAS и CMS

$$L_{comb}(\vec{\theta}) = L_{CMS}(\vec{\theta})L_{ATLAS}(\vec{\theta}), \quad (239)$$

где $\vec{\theta}$ – неизвестные параметры модели.

Предположим в обоих экспериментах мы искали одну и ту же сигнатуру, скажем, количество димюонов с двумя дополнительными адронными струями и с инвариантной массой $M(jet_1 + jet_2 + \mu^+ + \mu^-) \geq M_0$. Предположим, что мы знаем фоны $\lambda_{b,CMS}$ и $\lambda_{b,ATLAS}$ и предположим, что ожидаемое число событий (λ_{tot} в распределении Пуассона)

$$\begin{aligned} \lambda_{tot,CMS} &= \lambda_{b,CMS} + \lambda_{s,CMS}, \\ \lambda_{tot,ATLAS} &= \lambda_{b,ATLAS} + \lambda_{s,ATLAS}, \\ \lambda_{s,ATLAS} &= k\lambda_{s,CMS}, \end{aligned}$$

где k -известное число, связанное с различиями в детекторах CMS и ATLAS¹⁹. Предположим мы не зафиксировали ни одного события, удовлетворяющего нашим условиям. Тогда при отдельной обработке событий на 95% уровне достоверности мы имели бы

$$\begin{aligned} \lambda_{s,CMS} &< 3.0, \\ \lambda_{s,ATLAS} &< 3.0. \end{aligned}$$

Тогда как метод правдоподобия, как и метод комбинирования различных распределений Пуассона, что в данном случае одно и то же, дает

$$\lambda_{s,CMS} + \lambda_{s,ATLAS} < 3.0$$

или

$$\lambda_{s,CMS} < \frac{3.0}{1+k}.$$

Рассмотрим комбинирование двух распределений Пуассона $P(n_1, \lambda_1)$ и $P(n_2, \tau\lambda_1)$, где τ -фиксированное число. Логарифм функции правдоподобия

$$\ln L(\lambda_1) = n_1 \ln \lambda_1 + n_2 \ln(\tau\lambda_1) - \lambda_1 - \tau\lambda_1 - \ln(n_1!) - \ln(n_2!) \quad (240)$$

уравнения для определения максимума $\frac{\partial}{\partial \lambda_1}(\ln L) = 0$ приводит к

$$\lambda_1^{max} + \lambda_1^{max} \tau = n_1 + n_2,$$

¹⁹В случае тождественности CMS и ATLAS детекторов мы имели бы $k = 1$ и $\lambda_{b,CMS} = \lambda_{b,ATLAS}$.

$$\ln L(\lambda_1) - \ln L(\lambda_1 = \lambda_1^{max}) = (n_1 + n_2) \ln \frac{\lambda_1(1 + \tau)}{n_1 + n_2} - \lambda_1(1 + \tau) + n_1 + n_2. \quad (241)$$

Отсюда при заданном λ мы можем определить значимость или р-вероятность.

Заметим, что абсолютно такой же результат получается, если учесть, что сумма двух процессов Пуассона есть процесс Пуассона с $n = n_1 + n_2$ и $\lambda = \lambda_1 + \lambda_2$ и далее уже использовать метод максимального правдоподобия для распределения $P(n_1 + n_2, \lambda_1 + \lambda_2)$.

В более общем случае, когда мы явно учитываем фоны и $\lambda_1 = \lambda_{1b} + \lambda$ и $\lambda_2 = \lambda_{2b} + \tau\lambda$, где $\lambda_{1b}, \lambda_{2b}, \tau$ – фиксированные величины, а λ – параметр (число сигнальных событий для первого процесса), метод максимального правдоподобия и метод объединения двух Пуассоновских процессов приводят, вообще говоря, к разным результатам. Действительно в методе максимального правдоподобия уравнение $\frac{\partial}{\partial \lambda} L = 0$ имеет решение

$$\lambda = -\lambda_{1b} + \frac{n_1 + n_2}{2(1 + \tau)} - \frac{1}{2} \left(\frac{\lambda_{2b}}{\tau} - \lambda_{1b} \right) + \sqrt{\left(\frac{n_1 + n_2}{2(1 + \tau)} - \frac{1}{2} \left(\frac{\lambda_{2b}}{\tau} - \lambda_{1b} \right) \right)^2 - \frac{\lambda_{1b}\lambda_{2b}}{\tau} + \frac{1}{1 + \tau} \left(\frac{n_1\lambda_{2b}}{\tau} + n_2\lambda_{1b} \right)}. \quad (242)$$

Тогда как при комбинировании двух распределений Пуассона параметр λ равен

$$\lambda = \frac{n_1 + n_2 - \lambda_{1b} - \lambda_{2b}}{1 + \tau}. \quad (243)$$

В случае $\frac{\lambda_{2b}}{\tau} = \lambda_{1b}$ два подхода совпадают.

Заметим, что когда мы измеряем $n_{obs,CMS}$ и $n_{obs,ATLAS}$ число событий и находим ограничение сверху на параметры распределений Пуассона $\lambda_{tot,CMS}, \lambda_{tot,ATLAS}$, где $\lambda_{tot} = \lambda_b + \lambda_s$, обобщение вышеизложенного подхода очевидно.

6.6 Байесовский подход

Для N независимых измерений с функциями распределения $F_k(x|\vec{\theta})$ функция правдоподобия $L(\vec{\theta}) = \prod_{k=1}^N F_k(x_k|\vec{\theta})$. В этом случае согласно Байесовскому подходу комбинированная плотность вероятности

$$f(\vec{\theta}) = \frac{L(\vec{\theta})\pi(\vec{\theta})}{\int d\vec{\theta} L(\vec{\theta})\pi(\vec{\theta})}. \quad (244)$$

Остановка за малым – выбрать правильный prior $\pi(\vec{\theta})$. Формула (244) задает способ комбинирования результатов в Байесовском подходе.

Рассмотрим задачу комбинирования двух результатов, полученных на основе Пуассоновской статистики. Для случая $\lambda_1 = \lambda$, $\lambda_2 = k\lambda$, $k = const$ имеем тождество

$$P(n_1, \lambda_1)P(n_2, \lambda_2) = \frac{n_1!n_2!}{(n_1 + n_2)!} \frac{k^{n_2}}{(1+k)^{n_1+n_2}} P(n_1 + n_2, \lambda_1 + \lambda_2).$$

Иными словами, произведение двух распределений Пуассона с точностью до численного множителя совпадает с распределением Пуассона с числом событий $n_1 + n_2$ и параметром $\lambda = \lambda_1 + \lambda_2$.

При выборе плоского priora $\pi(\lambda) = const$ задача комбинирования свелась к задаче извлечения параметров из распределения $P(n_1 + n_2, \lambda_1 + \lambda_2)$. Для модели, описывающей как вклад фона, так и вклад сигнала с

$$\lambda_1 = \lambda_{b1} + \lambda_{s1},$$

$$\lambda_2 = \lambda_{b2} + k\lambda_{s1}$$

в случае, когда $\lambda_{b2} = k\lambda_{b1}$, произведение распределений Пуассона так же сводится с точностью до численного множителя к распределению Пуассона с $n = n_1 + n_2$ и $\lambda = \lambda_1 + \lambda_2$. В случае, когда $\lambda_{b2} \neq k\lambda_{b1}$ задача не сводится к распределению единичного Пуассона. В любом случае, для извлечения информации из двух измерений нам необходимо знать функцию priora $\pi(\lambda)$.

При комбинировании двух процессов Пуассона мы можем извлечь информацию, рассмотрев случайную величину равную сумме случайных величин, описывающихся распределениями Пуассона $P(n_i, \lambda_i)$, $i = 1, 2$, и описывающуюся процессом Пуассона с параметрами $n = n_1 + n_2$, $\lambda = \lambda_1 + \lambda_2$. При этом извлекать ограничения на параметр λ_s можно из распределения Пуассона $P(n_1 + n_2, \lambda_{b1} + \lambda_{b2} + (1+k)\lambda_{s1})$.

Заметим при этом, что учет систематических эффектов, связанных с неточным знанием λ_{b1} , λ_{b2} , λ_{s1} , λ_{s2} принципиально ничем не отличается от соответствующего рассмотрения для единичного распределения Пуассона.

Интересно отметить, что для распределения Пуассона $P(n_k, \lambda_k)$, где $\lambda_k = c_k\tau$ и c_k -фиксированные числа, а τ -неизвестный параметр, справедливо равенство

$$\prod_{k=1}^N P(n_k, \lambda_k) = cP(\sum n_k, \sum \lambda_k), \quad (245)$$

где коэффициент пропорциональности c зависит от величин n_k и c_k .

Равенство (245) является отражением того факта, что сумма процессов Пуассона является процессом Пуассона с наблюдаемым числом событий равным сумме наблюдаемых $n_{obs} = \sum n_i$ и параметром распределения λ равным сумме параметров распределения $\lambda = \sum \lambda_i$. На основании этой формулы задача комбинирования сводится к задаче определения параметра λ (параметра τ) для единичного процесса Пуассона.

6.7 Комбинирование уровней значимости (комбинирование значений вероятности)

Предположим, что мы хотим проверить основную гипотезу H_0 путем измерения N независимых величин. Предположим также, что у нас в результате измерений N независимых величин известны вероятности P_k их реализации или, что то же самое в силу формулы

$$\frac{1}{\sqrt{2\pi}} \int_S^\infty e^{-\frac{x^2}{2}} = P(S), \quad (246)$$

их значимостей S_k .

Встает вопрос: как определить комбинированную значимость или комбинированную вероятность. Существует несколько способов решения этой проблемы. В методе Фишера используется переменная $\alpha' = -2 \ln \prod_{i=1}^N P_i$. Далее предполагается, что α' распределена как $\chi^2(2N)$. Как следствие получаем формулу комбинирования. Для комбинирования двух вероятностей метод Фишера приводит к правилу комбинирования

$$P(P_1, P_2) = P_1 P_2 [1 - \ln(P_1 P_2)]. \quad (247)$$

В методе Стоуфера [31] был предложен другой способ комбинирования значимостей. Поясним этот метод на примере распределения Пуассона. Пусть мы измеряем N независимых величин, распределенных по закону Пуассона $P(n_k | \lambda_k)$ ²⁰ и пусть мы наблюдаем избыток событий в каждой из N независимых величин. Тогда мы можем определить вероятности

²⁰Предполагается, что параметры λ_k точно известны в рамках основной гипотезы.

$$P_k(n \geq n_{k,obs} | \lambda_k) = \sum_{n=n_{k,obs}}^{\infty} P(n | \lambda_k), \quad (248)$$

так называемые “p-value”.

В методе Стоуфера используется тот факт, что сумма процессов Пуассона является процессом Пуассона и, как следствие, формула комбинирования вероятностей имеет вид

$$P(P_1, \dots, P_n) = \sum_{n=\sum_k n_{k,obs}}^{\infty} P(n | \sum_k \lambda_k). \quad (249)$$

В случае $\lambda_k \ll n_{k,obs}$ сумма (249) определяется в основном первым членом в сумме

$$P_k(n \geq n_{k,obs} | \lambda_k) \approx \frac{1}{n_{k,obs}!} \lambda_k^{n_{k,obs}} e^{-\lambda_k} \quad (250)$$

и, следовательно,

$$P(P_1, \dots, P_n) = \frac{1}{(\sum_{k=1}^N n_{k,obs})!} (\sum_{k=1}^N \lambda_k)^{\sum_{k=1}^N n_{k,obs}} e^{-\sum_{k=1}^N \lambda_k}. \quad (251)$$

Для $n = 2$ сравнение правил комбинирования (247) и (251) показывает, что численные результаты могут сильно различаться.

В качестве примера рассмотрим комбинирование двух распределений Пуассона с $n_{1,obs} = n_{2,obs} = 1$ и $\lambda_1, \lambda_2 \ll 1$. В этом случае формулы (250) и (251) примут вид $P_1 \approx \lambda_1$, $P_2 \approx \lambda_2$,

$$P(P_1, P_2) = \frac{(\lambda_1 + \lambda_2)^2}{2} = \frac{(P_1 + P_2)^2}{2}. \quad (252)$$

Как видим, различие между формулой (252) и формулой Фишера (247) весьма существенно.

6.8 “Look elsewhere” эффект

Предположим, что мы ищем бозон Хиггса посредством реакции $pp \rightarrow H \rightarrow \gamma\gamma + \dots$, то есть мы ищем узкий пик в распределении дифотонных событий по инвариантной массе $m(\gamma\gamma)$. Предположим, что мы ищем эффект в интервале масс $m_- < m(\gamma\gamma) \leq m_+$. Стандартный путь – разбиваем интервал $[m_-, m_+]$ на k бинов, каждый с шириной $\Delta = \frac{m_+ - m_-}{k}$

и ищем усиление сигнала в одном из k бинов. Предположим для простоты, что распределение фона плоское, то есть фон не зависит от m при $m_- \leq m \leq m_+$.

Оценка фона в каждом из бинов в силу независимости распределения от m получается равной $\lambda = \frac{1}{k} \sum N_l$ и для большого количества k и N параметр фона λ определяется с хорошей точностью из экспериментальных данных.

Мы ищем превышение сигнала над фоном в каждом бине. Вероятность детектирования N_k событий в бине определяется формулой Пуассона (43) и является функцией от наблюдаемого числа частиц в бине и среднего λ ²¹.

“Look elsewhere” эффект состоит в том, что, поскольку мы не знаем точное расположение сигнала, вероятность реализации событий вследствие флуктуации фона увеличивается на фактор k по сравнению с наименее вероятным значением $P_{min} = \min_k P(N_{k,obs}|\lambda)$, где

$$P(N_{k,obs}|\lambda) = \sum_{n=N_{k,obs}}^{\infty} P(n|\lambda).$$

Иными словами “look elsewhere” эффект приводит к изменению вероятности $P_{min} \rightarrow k P_{min}$.

6.9 Тест Колмогорова-Смирнова

Предположим мы имеем набор измерений x_1, x_2, \dots, x_n и мы хотим проверить насколько совместим этот набор с функцией распределения $f(x)$.

В методе Колмогорова-Смирнова это достигается путем сравнения теоретической кумулятивной функции распределения

$$F(x) = \int_{-\infty}^x f(x') dx' \quad (253)$$

и “экспериментальной” кумулятивной функции распределения

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \theta(x - x_i), \quad (254)$$

где $\theta(x) = \begin{cases} 1, & x > 0, \\ 0, & x \leq 0. \end{cases}$

Расстояние между теоретической кумулятивной функцией распределения и экспериментальной определяется как

²¹Мы здесь рассматриваем случай, когда фон плоский и параметр λ , описывающий фон в распределении Пуассона, известен.

$$D_n = \sup_x |F_n(x) - F(x)|. \quad (255)$$

Для больших n функция $D_n \rightarrow 0$. Параметр $k = \sqrt{n}D_n$ имеет распределение $P(k \leq x) = 1 - 2 \sum_{l=1}^{\infty} (-1)^{l-1} e^{-l^2 x^2}$.

Подробное рассмотрение и доказательства можно найти в монографии [6].

С помощью теста Колмогорова-Смирнова можно решить вопрос следуют ли выборки x_1, \dots, x_n и y_1, \dots, y_N одному и тому же распределению. Для этого расстояние между двумя выборками определяется как

$$D_{n,m} = \sup_x |F_n(x) - F_m(x)|. \quad (256)$$

Можно показать, что переменная $\sqrt{\frac{nm}{n+m}} D_{n,m}$ асимптотически удовлетворяет распределению Колмогорова.

7 Статистическое программное обеспечение в задачах физики высоких энергий

7.1 Обзор основных пакетов

Базовым инструментарием для решения задач физики высоких энергий является система программ ROOT [5]. ROOT – пакет объектно-ориентированных программ и библиотек, разработанный в ЦЕРН. Проект базируется на свободном программном обеспечении. Наряду со специальными средствами программирования и стандартных математических вычислений, ROOT обеспечивает пользователя средствами для построения и анализа гистограмм и графиков функций, средствами фитирования и подбора теоретических и экспериментальных зависимостей, инструментарием для проведения статистического (в том числе многофакторного) анализа данных [32].

Большинство программных наработок, по возможности, либо встраиваются в ROOT, либо являются надстройкой над ROOT, то есть базируются на ROOTовских библиотеках. Проводятся работы по совмещению и/или по созданию интерфейса между пакетом ROOT и астрофизическим языком и оболочкой для статистических расчетов и построения графиков R [33].

Надстройкой над ROOT является пакет программ RooFit [34]. RooFit это инструментарий для моделирования ожидаемых распределений событий в физическом анализе. Пакет был изначально ориентирован на эксперимент ВаВаг. Пакет удобен для быстрого Монте Карло розыгрыша событий и статистической обработки полученных распределений. Со временем он стал универсальным и базовым для пакетов, расширяющих его возможности, например, RooStats [4].

Байесовская парадигма реализована в пакете программ BAT - The Bayesian Analysis Toolkit [35]. Анализ базируется на теореме Байеса и использует Монте Карло моделирование Марковских цепей. Это позволяет строить апостериорное распределение вероятностей, производить оценивание параметров, строить доверительные интервалы, осуществлять перенос неопределенностей.

Много внимания Статистические группы экспериментов уделяют статистическим пакетам многомерного анализа. Было проведено несколько мини-совещаний по данной тематике, например, Miniworkshop on Statistical Tools (2008) DESY [36], CMS tutorials on Multivariate Analysis Methods (2007) CERN [37]. Здесь можно выделить пакет TMVA - Toolkit for MultiVariate Data Analysis [38]. Данный инструментарий также встроен

в среду ROOT и ориентирован на использование многомерных классификационных алгоритмов для решения широкого спектра задач. Для решения задач физики высоких энергий и астрофизики также разрабатывается пакет StatPatternRecognition (SPR) [39]. Оба пакета имеют как перекрывающиеся, так и дополняющие друг друга возможности. Интересной новой разработкой является система поддержки принятия решений при выборе переменных в многомерном анализе и уменьшения размерности задачи PARADIGM [40].

7.2 Проект RooStats

Для анализа экспериментальных данных коллабораций CMS и ATLAS разрабатывается проект RooStats [4], основанный на комплексе программ ROOT [5]. Главные цели проекта

- предоставить пользователю компьютерные программы с наиболее распространенными статистическими методами, которые применяются при анализе данных в физике высоких энергий,
- стандартизовать используемые методы для легкого сравнения результатов полученных разными группами и разными экспериментами.

RooStats использует три наиболее распространенных подхода в статистике:

1. частотный подход,
2. метод максимального правдоподобия,
3. Байесовский подход.

Заметим, что программа *RooStats* постоянно развивается, совершенствуется и, на сегодняшний день, она содержит программы, позволяющие решать следующие задачи:

1. Точечная оценка для определения наилучшего в некотором смысле (например оценки с минимальной дисперсией или наиболее вероятные значения) значения параметра.
2. Определение доверительного интервала: областей параметров функции распределения, не противоречащих наблюдаемым данным.

3. Проверка гипотез: оценка значения вероятности p для одной или нескольких гипотез (значимость).
4. Оценка качества фита – количественное определение насколько хорошо модель описывает данные.

Программа RooStats написана на языке C++ и содержит следующие классы, позволяющие решать эти задачи.

- *ProfileLikelihoodCalculator* вычисляет значимость сигнала и определяет наилучшее значение сигнала на основе метода максимального правдоподобия. Возможность учета систематических эффектов также включена в калькулятор.
- *ProfileLikelihood* возможно использовать для оценки интервалов доверия. Возможно вычисление верхнего и нижнего пределов, а также центрального доверительного интервала.
- *BayesianCalculator* позволяет решать задачи на основе метода Байеса. Причем Байесовское интегрирование может производиться численно, аналитически и методом Монте Карло с помощью Марковских цепей (*MCMCCalculator* - Monte Carlo Markov Chain Calculator). Здесь, конечно, очень важно – выбор функции приора $\pi(\lambda)$, а также выбор интервала – центрального интервала, интервала минимальной длины или одностороннего интервала.
- *HybridCalculator* вычисляет частотную вероятность событий. Учет систематики проводится с помощью метода Кузинса-Хайлэнда. В частности вычисляет “п”-значения (p-value). Вычисления производятся с помощью Монте Карло розыгрыша псевдоэкспериментов. Также возможно осуществить построение Неймана (*NeymanConstruction*) для определения интервалов доверия частотным способом. Предоставляется возможность использовать несколько правил конструирования интервалов (интервал минимальной длины, центральный интервал, метод Фельдмана-Кузинса). Также при определении значения верхних пределов на сигнал можно использовать чисто P_{sl} частотный подход и $CL_s = \frac{P_{sl}}{1 - p_B}$ – модифицированный частотный подход.
- *HypoTestInverter* преобразует результат по проверке гипотез (*HybridCalculator*) в доверительный интервал (или в предел доверия) для параметра.

- *HistFactory* обеспечивает использование статистического инструментария пакета программ RooStats без необходимости использовать язык моделирования данных для пакета RooFit.
- *BATCalculator*. Сам пакет Байесовских вычислений с помощью Монте Карло цепей Маркова (BAT) является внешним к RooStats, но как класс в пакете программ RooStats полезен.

7.3 Проект BAT

BAT (the Bayesian Analysis Toolkit) [41] – инструментарий для статистического анализа также как и RooStats возник недавно. Статистический анализ данных в пакете программ BAT основывается на теореме Байеса и реализуется с помощью метода Монте Карло с использованием цепей Маркова [42]. Это позволяет строить апостериорные распределения параметров и, соответственно, проводить оценку параметров, устанавливать доверительные пределы и интервалы, а также осуществлять непосредственно перенос неопределенностей на уровне апостериорных распределений.

Одна из главных целей анализа данных – сравнить модельные предсказания с экспериментальными данными и либо сделать заключение о корректности модели по отношению к данным, либо построить области доверия с той или иной точностью для параметров заданной модели.

Заключение

Основной задачей Большого адронного коллайдера является поиск бозона Хиггса – последней неоткрытой частицы, предсказываемой Стандартной Моделью. Так же весьма важен поиск проявлений новой физики вне рамок Стандартной Модели, включая в первую очередь суперсимметрию. При обработке экспериментальных данных и получении ограничений на новые частицы и взаимодействия используются частотный подход, метод максимального правдоподобия и Байесовский подход. Как правило эти подходы приводят к численно близким результатам. При извлечении ограничений на новую физику из данных БАКа очень важно правильно оценивать влияние систематических эффектов, что во многих случаях весьма и весьма нетривиально.

Авторы благодарны академику В.А. Матвееву по инициативе которого был написан этот обзор. Работа поддержана грантом РФФИ 10-02-00468-а.

Список литературы

- [1] CMS, Technical Proposal, CERN/LHCC/94-38.
- [2] ATLAS, Technical Proposal, CERN/LHCC/94-43.
- [3] T.L.S.Group: **The Large Hadron Collider Conceptual Design**. CERN-AC-95-05 (1995); arXiv:hep-ph/061012.
- [4] <https://twiki.cern.ch/twiki/bin/view/RooStats/>
- [5] R. Brun, F. Rademaker, Nucl.Instr.&Meth., **A389** (1997) 81; см., также, <http://root.cern.ch/>
- [6] W.T.Eadie, D.Drijard, F.E.James, M.Roos, and B.Sadoulet, *Statistical Methods in Experimental Physics*, North Holland, Amsterdam, 1971; русский перевод: *Статистические методы в экспериментальной физике*, АТОМИЗДАТ, Москва, 1976; F. James *Statistical methods in experimental physics*, 2nd ed., World Scientific, 2006.
- [7] A.G.Frodesen, O.Skjeggestad, H.Tøft, *Probability and Statistics in Particle Physics*, UNIVERSITETSFORLAGET, Bergen-Oslo-Tromsø, 1979.
- [8] Д. Худсон, *Статистика для физиков*, МИР, Москва, 1970.
L. Lyons, *Statistics for Particle and Nuclear Physicists*, Cambridge University Press, 1986.
G. Cowan, *Statistical Data Analysis*, Oxford Univ. Press, 1998.
R. J. Barlow, *Statistics*, J. Wiley, 1989.
D. Sivia, J. Skilling, *Data Analysis, A Bayesian Tutorial*, 2nd ed., Oxford University Press, 2006.
- [9] G. D’Agostini, *Bayesian Reasoning in Data Analysis, a Critical Introduction*. World Scientific, Hackensack, NJ, 2003.
- [10] Г. Крамер, *Математические методы статистики*, МИР, Москва, 1976.
- [11] Н.В. Красников, В.А. Матвеев, *Новая физика на Большом адронном коллайдере*, “Красанд”, Москва, 2011.
- [12] A.N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer, Berlin, 1933.

- [13] Bayes, Thomas, and Price, Richard, *An Essay towards solving a Problem in the Doctrine of Chances*, by the later Rev. Th. Bayes, communicated by Mr. Price, in a letter to John Canton, M. A. and F. R. S., Philosophical Transactions of the Royal Society of London, **53** (1763) 370-418.
- [14] E.L. Lehmann, G. Casella, *Theory of Point Estimation* (2nd ed.), Springer, 1998.
- [15] S. Bityukov, N. Krasnikov, S. Nadarajah, V. Smirnova, *Confidence distributions in statistical inference*. “Bayesian Inference and Maximum Entropy Methods in Science and Engineering”, 30-th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Chamonix, France, 4-9 July 2010, Eds: A. Mohammad-Djafari, J-F. Bercher, P. Bessiere, AIP Conference Proceedings, vol.**1305**:446-456, Melville, NY, 2010.
- [16] J. Neyman, *Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability*, Philosophical Transactions of the Royal Society of London A, **236** (1937) 333-380.
- [17] K. Nakamura et al. (Particle Data Group), *J. Phys.* **G 37**, 075021 (2010).
- [18] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*, (John Wiley and Sons, Chichester, UK, 1993);
H. Jeffreys, *Theory of Probability*, 3rd ed., (Oxford University Press, Oxford, 1961);
- [19] J. Heinrich, et al., *Interval estimation in the presence of nuisance parameters. 1. Bayesian approach*, CDF/MEMO/STATISTICS/PUBLIC/7117; arXiv:physics/0409129 [physics.data-an], 2004.
- [20] G. Casella, R.L. Berger, *Statistical Inference*, Duxbury Press, (1990) 346;
- [21] CMS Collaboration. **CMS Physics, Technical Design Report Volume II: Physics Performance**. CERN/LHCC 2006-021, CMS TDR 8.2.
- [22] G. Feldman, R. Cousins, *Phys.Rev.* **D57**, 3873-3889,1998

- [23] Helene O. (1988). G.P. Yost et.al., Rev.Part.Prop., *Phys.Lett.*, **B204**, p.81.
- [24] A.L. Read, *Presentation of search results: the CLs technique*, J.Phys.G: Nucl. Part. Phys. **28** (2002) 2693-2704.
- [25] The CMS Collaboration, *Measurement of CMS Luminosity*, CMS PAS EWK-10-004, 2010.
- [26] R.D. Cousins and V.L. Highland, *Incorporating systematic uncertainties into an upper limit*, Nucl.Instr.&Meth. **A320** (1992) 331-335.
- [27] S.L. Glashow, Nucl.Phys., **22** 579 (1961);
S. Weinberg, Phys.Rev.Lett., **19** 1264 (1967);
A. Salam, Elementary Particle Theory (ed. N.Svartholm), Almquist and Wiksells, Stockholm, 1964;
H.D. Politzer, Phys.Rev.Lett., **30** 1346 (1973);
D.J. Gross, F.E. Wilczek, Phys.Rev.Lett. **30** 1343 (1973).
- [28] Y. Fukuda et. al., Phys.Rev.Lett. **81** 1562 (1998);
K. Eguchi et. al., Phys.Rev.Lett. **90** 021802 (2003);
M.H. Ahn et al., Phys.Rev.Lett. **90** 041803 (2003)
- [29] Ю.А. Гольфанд, Е.П. Лихтман, Письма в ЖЭТФ, **13**, (1971) 452;
Д.В. Волков, В.П. Акулов, Письма в ЖЭТФ, **16** (1972) 621;
J. Wess, B. Zumino, Nucl.Phys., **B70** (1974) 39.
- [30] А.И. Кибзун и др., Теория вероятностей и математическая статистика, Москва, ФИЗМАТЛИТ, 2002.
- [31] S. Stouffer, E. Suchman, L. DeVinnery, S. Star, and R.W. Jr, *The American Soldier*, volume I: Adjustment during Army Life. Princeton University Press, 1949.
- [32] L. Moneta, I. Antcheva, R. Brun, A. Kreshuk, *ROOT Statistical Software*, Proceedings of PhyStat-LHC, CERN-2008-001, pp. 179-183.
- [33] <http://www.r-project.org/>
- [34] <http://roofit.sourceforge.net>

- [35] <http://www.mppmu.mpg.de/bat/>
- [36] <https://indico.desy.de/conferenceDisplay.py?confId=1097>
- [37] <http://indico.cern.ch/conferenceDisplay.py?confId=24781>
- [38] <http://tmva.sourceforge.net/>
- [39] <http://sourceforge.net/projects/statpatrec>
- [40] S.V. Gleyzer, H. Prosper, *PARADIGM, a Decision Making Framework for Variable Selection and Reduction in High Energy Physics*, Proceedings of Science (ACAT08) 067, 2008.
- [41] A. Caldwell, D. Kollar, K. Kroninger, *BAT- The Bayesian analysis toolkit*, Computer Physics Communications **180** (2009) 2197-2209.
- [42] Дж.Дж. Кемени, Дж.Л. Снелл, Конечные цепи Маркова, Москва, "Наука", 1970.