

Качество согласия и статистическая значимость

Содержание

13.1.	Физическая задача: можно ли пользоваться калибровкой?	176
13.2.	Остатки и статистика качества согласия	176
13.3.	Откуда берётся распределение χ^2	177
13.4.	Сколько степеней свободы осталось после фита?	179
13.5.	p-значение: какую вероятность мы вычисляем?	180
13.6.	Возвращаемся к двум калибровкам	182
13.7.	Что означает «число сигма»?	183
13.8.	Читаем графики поиска бозона Хиггса	185
13.9.	Когда распределение нужно получить моделированием	187
13.10.	Задачи	189
	Литература	191

Аннотация

Подгонка даёт параметры модели и их ошибки. В этой главе проверим, насколько сама модель согласуется с данными. Выведем распределение χ^2 , разберём число степеней свободы и научимся вычислять р-значение. Вернёмся к графикам поиска бозона Хиггса из первой главы и выясним, что означают линии «числа сигма». Для малых пуассоновских счётов получим распределение статистики псевдоэкспериментами.

13.1. Физическая задача: можно ли пользоваться калибровкой?

Мы откалибровали детектор по источникам с известными энергиями. В каждой точке измерили среднюю амплитуду сигнала и подогнали прямую

$$y_i = a + bE_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_i^2). \quad (13.1)$$

Здесь E_i — энергия в МэВ, y_i — амплитуда в каналах, a — смещение нуля, b — коэффициент преобразования в каналах на МэВ. Энергии считаем известными точно, ошибки амплитуд — независимыми и гауссовыми, а σ_i — известными до подгонки.

Программа нашла $\hat{a}$, $\hat{b}$ и их ковариацию. Можно теперь пересчитывать амплитуды в энергии? Только если прямая описывает отклик детектора с нужной нам точностью. Электроника может давать нелинейность; при больших амплитудах возможно насыщение. Минимизатор об этом ничего не знает. Он найдёт наиболее подходящую прямую для тех точек, которые ему передали.

Возьмём два учебных результата. В первой калибровке было $N = 12$ точек, и после подгонки двух параметров получилось $\chi_{\min}^2 = 12$. Во второй — $N = 1002$ точки и $\chi_{\min}^2 = 1200$. Числа степеней свободы равны соответственно 10 и 1000. В обоих случаях

$$\frac{\chi_{\min}^2}{\nu} = 1.2. \quad (13.2)$$

В физическом фольклоре отношение, близкое к единице, часто принимают за достаточную проверку фита. Наши два примера показывают, чего в такой проверке не хватает. При верной модели вероятность получить значение $\chi_{\min}^2$ не меньше указанного равна примерно 0.285 в первом случае и $1.23 \cdot 10^{-5}$ во втором.

Разница больше четырёх порядков. Давайте получим эти числа и разберёмся, что именно они позволяют сказать о калибровке.

13.2. Остатки и статистика качества согласия

После подгонки остатки равны

$$r_i = y_i - \hat{a} - \hat{b}E_i, \quad u_i = \frac{r_i}{\sigma_i}, \quad \chi_{\min}^2 = \sum_{i=1}^N u_i^2. \quad (13.3)$$

Остаток r_i измеряется в каналах, нормированный остаток u_i безразмерен. Точка с $u_i = 3$ находится на расстоянии трёх своих исходных ошибок от подогнанной прямой и вносит в сумму квадратов девятку.

Но сами u_i после подгонки уже зависимы. Для прямой нормальные уравнения требуют

$$\sum_i \frac{r_i}{\sigma_i^2} = 0, \quad \sum_i \frac{E_i r_i}{\sigma_i^2} = 0. \quad (13.4)$$

Мы использовали эти же точки, чтобы выбрать прямую. Поэтому остатки связаны двумя условиями, а их дисперсии меньше дисперсий исходных ошибок. Нельзя считать все u_i независимыми величинами $\mathcal{N}(0, 1)$. В главе 11 это было видно из матрицы проекции.

Величина $\chi_{\min}^2$ — **статистика**, то есть вычисляемая по данным числовая функция. При повторении калибровки изменятся амплитуды, оценки $\hat{a}$, $\hat{b}$ и минимум суммы квадратов. Полученное число нужно сравнить с распределением этой статистики в повторяемых опытах при верной модели [1].

При этом $\chi_{\min}^2$ и $\Delta\chi^2$ из прошлой главы отвечают на разные вопросы. Минимум характеризует остаточное расхождение данных с моделью после подгонки. Приращение относительно минимума показывает, как меняется согласие при смещении параметров. Распределения у этих двух величин тоже разные.

13.3. Откуда берётся распределение χ^2

13.3.1. Квадрат одной гауссовой величины

Начнём со стандартной гауссовой величины:

$$G \sim \mathcal{N}(0, 1), \quad \varphi(g) = \frac{1}{\sqrt{2\pi}} e^{-g^2/2}, \quad T = G^2. \quad (13.5)$$

Букву Z оставим для значимости, которая появится ниже. Для $t > 0$ уравнение $g^2 = t$ имеет два корня: $\sqrt{t}$ и $-\sqrt{t}$. При замене переменной нужно учесть оба, как мы делали в главе 5.

Вывод

Для каждой ветви $|dg/dt| = 1/(2\sqrt{t})$. Поэтому

$$\begin{aligned} f_T(t) &= \frac{\varphi(\sqrt{t}) + \varphi(-\sqrt{t})}{2\sqrt{t}} \\ &= \frac{1}{\sqrt{2\pi t}} e^{-t/2}, \quad t > 0. \end{aligned} \quad (13.6)$$

Это плотность распределения χ^2 с **одной степенью свободы**. Отрицательных значений у T нет. При приближении к нулю плотность растёт без границы, но её интеграл остаётся конечным. Бесконечная плотность в одной точке не означает бесконечной вероятности.

13.3.2. Сумма квадратов и геометрия

Теперь возьмём v независимых величин $G_i \sim \mathcal{N}(0, 1)$:

$$T = \sum_{i=1}^v G_i^2, \quad T \sim \chi_v^2. \quad (13.7)$$

Их совместная плотность пропорциональна $\exp(-\sum_i g_i^2/2)$. Она зависит только от расстояния до начала координат. В двумерном случае мы уже интегрировали такую плотность по кругам; теперь число измерений произвольно.

Вывод

Для сферического слоя радиуса r вероятность пропорциональна

$$e^{-r^2/2} r^{v-1} dr.$$

Здесь $r^{v-1} dr$ даёт зависимость объёма слоя от радиуса. При переходе к $t = r^2$ получаем $r^{v-1} dr = \frac{1}{2} t^{v/2-1} dt$. Нормировка приводит к плотности

$$f(t; v) = \frac{t^{v/2-1} e^{-t/2}}{2^{v/2} \Gamma(v/2)}, \quad t > 0. \quad (13.8)$$

Гамма-функция определяется интегралом

$$\Gamma(a) = \int_0^\infty u^{a-1} e^{-u} du, \quad a > 0. \quad (13.9)$$

Из интегрирования по частям следует $\Gamma(a+1) = a\Gamma(a)$. В частности, $\Gamma(n) = (n-1)!$ для целого $n \geq 1$ и $\Gamma(1/2) = \sqrt{\pi}$. Подстановка $v = 1$ в 13.8 возвращает результат 13.6.

13.3.3. Среднее, ширина и число степеней свободы

Найдём первые два момента. В обоих интегралах достаточно сделать замену $t = 2u$:

$$\begin{aligned} \mathbb{E}[T] &= 2 \frac{\Gamma(v/2 + 1)}{\Gamma(v/2)} = v, \\ \mathbb{E}[T^2] &= 4 \frac{\Gamma(v/2 + 2)}{\Gamma(v/2)} = v(v+2). \end{aligned} \quad (13.10)$$

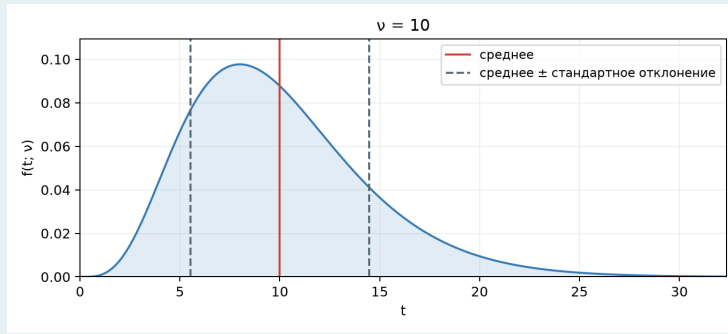
Отсюда

$$\text{Var}(T) = 2v, \quad \mathbb{E}[T/v] = 1, \quad \sigma_{T/v} = \sqrt{\frac{2}{v}}. \quad (13.11)$$

Последняя формула объясняет проблему с правилом « χ^2 на степень свободы около единицы». При $v = 10$ стандартное отклонение отношения равно 0.447, при $v = 1000$ — 0.0447. То же превышение единицы на 0.2 во втором случае в десять раз больше в единицах ширины распределения.

13.3.4. Интерактив: форма распределения

Меняйте число степеней свободы v . Сплошная вертикальная линия показывает среднее, пунктир — расстояние $\sqrt{2v}$ от него. При $v = 1$ верхушка плотности у нуля выходит за пределы рисунка; сама формула остаётся той же.



Интерактив. Форма распределения χ^2

При $\nu = 10$ среднее равно 10, стандартное отклонение — $\sqrt{20}$. Плотность асимметрична. Отметки среднего и ширины не являются границами интервала с заданной вероятностью.



С ростом ν распределение становится более симметричным. В центральной области работает гауссово приближение $(T - \nu)/\sqrt{2\nu} \approx \mathcal{N}(0, 1)$. Для далёкого хвоста точность этого приближения нужно проверять: именно там нас могут интересовать вероятности порядка 10^{-5} и меньше.

13.4. Сколько степеней свободы осталось после фита?

13.4.1. Коррелированные ошибки

Пусть данные гауссовы, их средние λ заданы заранее, а ковариация V_y известна и положительно определена:

$$\chi^2 = (\mathbf{y} - \lambda)^T V_y^{-1} (\mathbf{y} - \lambda). \quad (13.12)$$

Между измерениями могут быть сильные корреляции. Уменьшилось ли из-за этого число независимых направлений? Проверим заменой координат.

Диагонализуем ковариацию и нормируем каждую координату на её ошибку:

$$\begin{aligned} V_y &= Q D Q^T, \quad D = \text{diag}(d_1, \dots, d_N), \\ \mathbf{g} &= D^{-1/2} Q^T (\mathbf{y} - \lambda). \end{aligned} \quad (13.13)$$

Матрица Q ортогональна, все $d_i > 0$. В новых координатах ковариация единичная. Для совместно гауссовых данных это означает независимость компонент, и

$$\chi^2 = \sum_{i=1}^N g_i^2 \sim \chi_N^2. \quad (13.14)$$

Осталось ровно N направлений. Корреляции изменили их ориентацию и масштабы. При вырожденной ковариации появляются точные ограничения: тогда сначала выделяют допустимое подпространство. Обычной обратной матрицы в формуле 13.12 уже нет.

13.4.2. Подгонка выбирает проекцию

Для линейной модели $X\theta$ с t независимыми параметрами предсказания образуют t -мерное линейное подпространство. Здесь t — ранг матрицы модели; два одинаковых столбца не дают двух независимых параметров.

После преобразования 13.13 подгонка выбирает ближайшую точку подпространства модели. Остаток ортогонален ему и имеет $N - m$ независимых компонент. Если модель верна, все эти компоненты стандартные гауссовы. Поэтому [1]

$$\chi_{\min}^2 \sim \chi_v^2, \quad v = N - m. \quad (13.15)$$

Это точный результат для линейной гауссовой задачи с известной ковариацией и свободными параметрами. Если истинный параметр лежит на физической границе или модель существенно нелинейна, такая геометрия может перестать работать.

13.4.3. Куда уходит одна степень свободы

Для измерений одной величины с одинаковыми известными ошибками

$$Y_i \sim \mathcal{N}(\mu, \sigma^2), \quad \hat{\mu} = \bar{Y}. \quad (13.16)$$

До подгонки сумма $\sum_i (Y_i - \mu)^2 / \sigma^2$ имеет N степеней свободы. После подгонки $\sum_i (Y_i - \bar{Y}) = 0$: свободных остатков стало на один меньше.

Это можно записать точным равенством

$$\sum_i \frac{(Y_i - \mu)^2}{\sigma^2} = \sum_i \frac{(Y_i - \bar{Y})^2}{\sigma^2} + \frac{N(\bar{Y} - \mu)^2}{\sigma^2}. \quad (13.17)$$

Последнее слагаемое — квадрат одной стандартной гауссовой величины. В гауссовой модели оно независимо от первой суммы справа. Подгонка убрала одно направление, оставив χ_{N-1}^2 .

У калибровочной прямой два независимых параметра, если энергии источников не все одинаковы. Поэтому в наших двух калибровках $v = 12 - 2 = 10$ и $v = 1002 - 2 = 1000$.

13.4.4. А если есть калибровочные ограничения?

В прошлой главе мы включали в правдоподобие вспомогательные измерения. Их нужно учитывать и при проверке качества согласия. В линейной гауссовой задаче с N основными и K независимыми вспомогательными данными после подгонки r независимых параметров остаётся $N + K - r$ степеней свободы.

Например, общая неизвестная поправка калибровки добавляет один параметр, но независимое измерение этой поправки добавляет одно наблюдение. Просто вычесть параметр из прежнего N было бы неверно. Этот подсчёт предполагает, что в повторяемом опыте флуктуируют и основные, и вспомогательные данные. Для другой процедуры распределение минимума проверяют отдельно.

13.5. p -значение: какую вероятность мы вычисляем?

Обозначим проверяемую гипотезу через H_0 . В калибровочном примере она утверждает, что отклик линейный, ошибки гауссовы и их ковариация задана верно. Параметры прямой подгоняются по данным.

Пусть получено $t_{\text{obs}} = \chi_{\min}^2$. Сумма квадратов тем больше, чем сильнее остаточное расхождение. Поэтому берём правый хвост:

$$p = P_{H_0}(T \geq t_{\text{obs}}) = \int_{t_{\text{obs}}}^{\infty} f(t; v) dt. \quad (13.18)$$

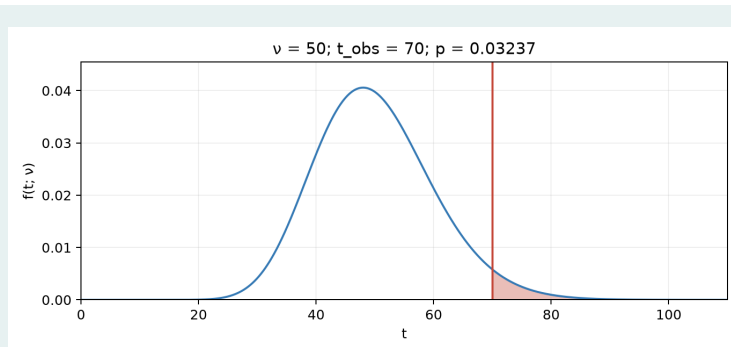
Спросим словами: если повторять эксперимент при верной H_0 , в какой доле опытов получится сумма квадратов хотя бы такого размера?

Вопрос о вероятности самой гипотезы здесь не ставился. Поэтому из $p = 0.03$ нельзя получить утверждение «вероятность модели равна 3%» или «модель неверна с вероятностью 97%». Число относится к распределению возможных результатов при заданной гипотезе.

Если нужен формальный критерий отклонения, заранее выбирают **уровень значимости** α и отклоняют H_0 при $p \leq \alpha$. Для точной непрерывной калибровки доля ложных отклонений верной гипотезы будет равна α . Значение α задаётся правилом проверки, а p вычисляется по конкретным данным.

13.5.1. Интерактив: площадь хвоста

В апплете можно менять v и полученное значение t_{obs} . При движении вертикальной линии вправо закрашенная площадь уменьшается. Вероятность вычисляется по функции распределения; точность числа не зависит от густоты точек на графике.



Интерактив. р-значение как площадь хвоста

Для $v = 50$ и $t_{\text{obs}} = 70$ площадь правого хвоста равна $p \approx 0.0324$. Закрашена вероятность получить $T \geq 70$ при этой модели.



Большое p означает, что выбранная статистика не обнаружила сильного несогласия. Оно не устанавливает истинность модели: другая модель тоже может описывать эти данные, а статистики может не хватить для различения моделей.

Очень маленькое χ^2 тоже даёт повод проверить расчёт. Возможно, ошибки завышены или часть данных была учтена несколько раз как независимая. Но иногда малое значение возникает просто из-за флуктуации. Правохвостовое p , близкое к единице, само по себе не является критерием для этой отдельной проверки.

13.5.2. Как распределены сами р-значения?

Представим много опытов, в которых H_0 верна. Для каждого вычислим T и его p -значение. Если функция распределения F_0 непрерывна, то $p = 1 - F_0(T)$. Преобразование переменной даёт

$$P_{H_0}(p \leq u) = u, \quad 0 \leq u \leq 1. \quad (13.19)$$

Плотность p -значений равна единице на отрезке $[0, 1]$. Значения около 0.5 ничем не выделены. При пороге $\alpha = 0.05$ примерно каждый двадцатый опыт с верной моделью попадёт в область отклонения.

В апплете одна выборка из χ^2_{10} задаёт численную калибровку, а независимая вторая выборка изображает повторные эксперименты. Число опытов во второй выборке можно менять. Для каждого результата подсчитывается правый хвост

калибровочной выборки.



Интерактив. Распределение p-значений

20 000 опытов задают калибровку; для 5000 независимых опытов построена гистограмма p-значений. Пунктир показывает единичную плотность. Отклонения столбцов связаны с конечным размером обеих выборок.



При конечной калибровочной выборке равномерность лишь приближённая. Для дискретной статистики точное распределение p-значений также ступенчатое. При определении через $P(T \geq t_{\text{obs}})$ получается консервативное правило:

$$P_{H_0}(p \leq \alpha) \leq \alpha. \quad (13.20)$$

Например, при малом пуассоновском счёте может вообще не существовать порога, дающего ровно 5% ложных отклонений. Это свойство дискретных данных, которое следует учитывать при чтении гистограммы p-значений.

13.6. Возвращаемся к двум калибровкам

Теперь вычислим интеграл 13.18 для результатов из начала главы. Для сравнения приведём и отклонение от среднего в единицах ширины:

$$z_{\chi^2} = \frac{t_{\text{obs}} - \nu}{\sqrt{2\nu}}. \quad (13.21)$$

Таблица 13.1. Два результата калибровки с одинаковым отношением χ^2/ν

t_{obs}	ν	t_{obs}/ν	z_{χ^2}	p
12	10	1.20	0.447	0.2851
1200	1000	1.20	4.472	$1.226 \cdot 10^{-5}$

Первая калибровка не вызывает возражений по этой проверке. Во второй остатки слишком велики для принятой модели ошибок. Нужно смотреть, где именно возникло расхождение: кривизна отклика, отдельная точка, зависимость ошибки от энергии, забытая общая систематика.

Само p-значение не выбирает причину. Оно относится ко всей проверяемой модели, включая ошибки. Если увеличить все σ_i , то χ^2 уменьшится. Но подбирать ошибки так, чтобы отношение стало единицей, — значит менять исходную постановку по результату этой же проверки.

Поэтому запись результата должна сохранять оба числа:

$$\chi_{\min}^2/\text{ndf} = 1200/1000, \quad p = 1.226 \cdot 10^{-5}. \quad (13.22)$$

Сокращение ndf означает число степеней свободы. Одна запись « $\chi^2/\text{ndf} = 1.2$ » теряет информацию, от которой зависит вывод.

13.7. Что означает «число сигма»?

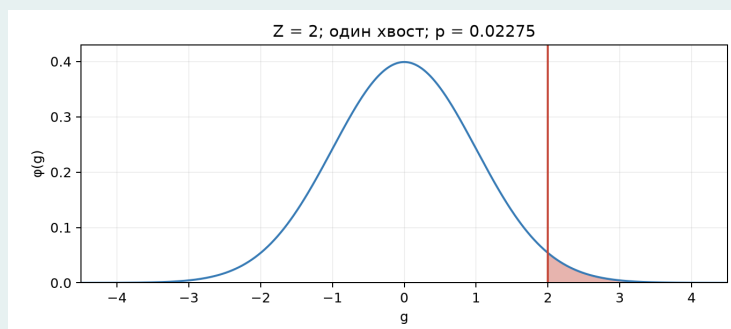
13.7.1. Один хвост

Малую хвостовую вероятность удобно перевести на знакомую гауссову шкалу. Для стандартной гауссовой величины G определим число Z условием

$$p = P(G \geq Z) = 1 - \Phi(Z), \quad Z = \Phi^{-1}(1 - p). \quad (13.23)$$

Функция Φ — функция распределения $\mathcal{N}(0, 1)$. Это соглашение о представлении вероятности. Исходные данные могут быть пуассоновскими, а статистика — иметь распределение χ^2 . Гауссовой здесь служит шкала, на которую мы переводим уже найденное p .

Для второго результата в таблице 13.1 односторонний перевод даёт $Z \approx 4.22$. Число $z_{\chi^2} = 4.472$ в той же таблице другое: оно получено вычитанием среднего и делением на стандартное отклонение. У асимметричного распределения эта операция не воспроизводит точную вероятность хвоста.



Интерактив. Односторонняя гауссова шкала

При $Z = 2$ правый хвост содержит $p = 0.02275$. Ползунок в HTML позволяет сопоставить положение границы с площадью хвоста.



Таблица 13.2. Одностороннее соответствие p и Z

Z	$p = 1 - \Phi(Z)$
1	$1.587 \cdot 10^{-1}$
2	$2.275 \cdot 10^{-2}$
3	$1.350 \cdot 10^{-3}$
4	$3.167 \cdot 10^{-5}$
5	$2.867 \cdot 10^{-7}$

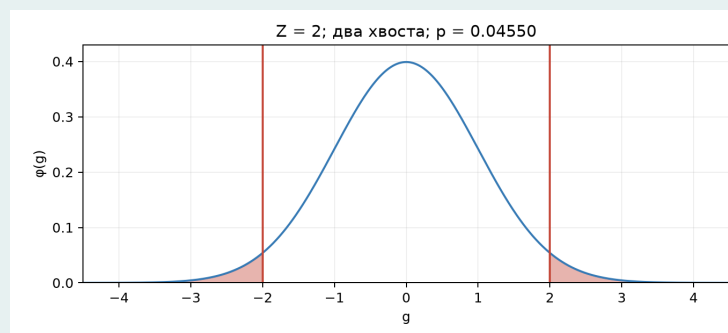
При поиске положительного сигнала обычно используют именно такое соглашение. Порог открытия 5σ соответствует правому хвосту из таблицы 13.2. Если $p > 1/2$, формула 13.23 даёт отрицательное Z . Такое число не означает свидетельства в пользу положительного сигнала.

13.7.2. Два хвоста

Если нас интересуют отклонения от заданного значения в обе стороны, складываются две хвостовые вероятности:

$$p = P(|G| \geq Z) = 2[1 - \Phi(Z)], \quad Z = \Phi^{-1}(1 - p/2), \quad Z \geq 0. \quad (13.24)$$

Для $Z = 2$ теперь получается $p = 0.04550$, вдвое больше одностороннего значения. Выбор определяется вопросом анализа. Его нельзя менять после того, как стал известен знак отклонения.



Интерактив. Два гауссовых хвоста

За пределами $[-2, 2]$ находится вероятность 0.04550. Каждый из двух хвостов содержит половину этой вероятности.



Уже знакомый интервал для одного параметра в квадратичном приближении строится из

$$\Delta\chi^2(\theta) \approx \frac{(\theta - \hat{\theta})^2}{\sigma_{\hat{\theta}}^2}. \quad (13.25)$$

Его границы $\hat{\theta} \pm Z\sigma_{\hat{\theta}}$ соответствуют $\Delta\chi^2 = Z^2$. В точной линейной гауссовой задаче покрытие такого интервала равно $2\Phi(Z) - 1$.

Таблица 13.3. Двусторонние гауссовы интервалы одного параметра

Границы интервала	Вероятность двух хвостов	$\Delta\chi^2$ на границе
$\hat{\theta} \pm \sigma_{\hat{\theta}}$	$3.173 \cdot 10^{-1}$	1
$\hat{\theta} \pm 2\sigma_{\hat{\theta}}$	$4.550 \cdot 10^{-2}$	4
$\hat{\theta} \pm 3\sigma_{\hat{\theta}}$	$2.700 \cdot 10^{-3}$	9
$\hat{\theta} \pm 4\sigma_{\hat{\theta}}$	$6.334 \cdot 10^{-5}$	16
$\hat{\theta} \pm 5\sigma_{\hat{\theta}}$	$5.733 \cdot 10^{-7}$	25

Это утверждение относится к приращению функции для одного параметра. Подставлять вместо него полное $\chi_{\min}^2$ нельзя.

13.8. Читаем графики поиска бозона Хиггса

Вернёмся к рисункам из первой главы. По горизонтали стоит пробная масса m_H , по вертикали — локальное p -значение фоновой гипотезы. В каждой точке спрашивают: как часто один фон даёт избыток, настолько же похожий на сигнал с этой массой?

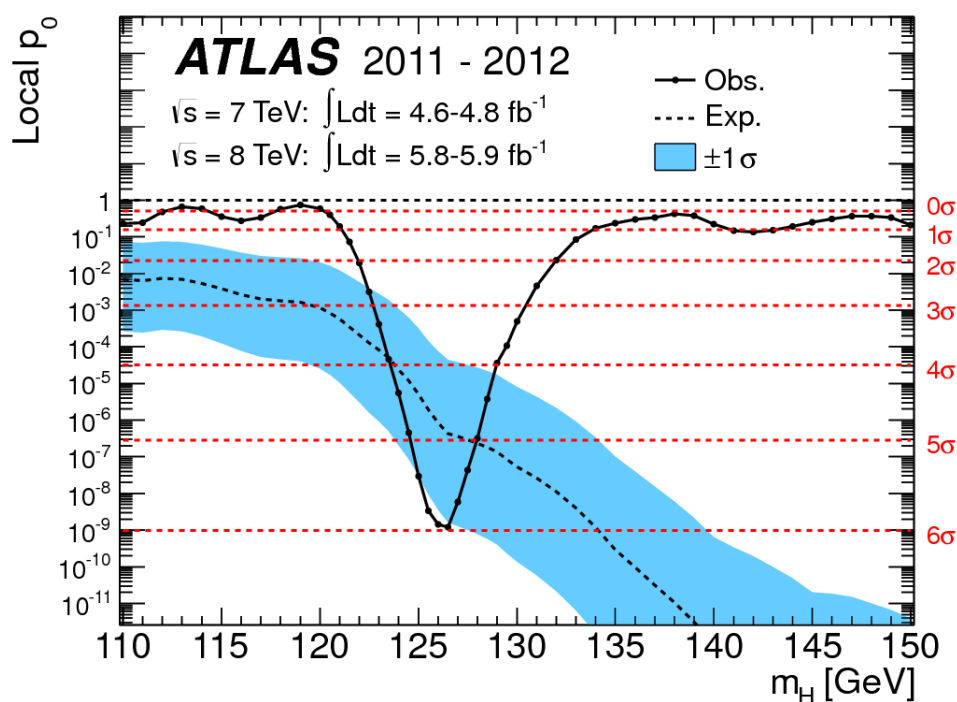


Рис. 13.1. Локальное p -значение фоновой гипотезы в ATLAS. Сплошная чёрная кривая получена из данных; пунктир и голубая полоса относятся к ожидаемому результату при наличии бозона Хиггса Стандартной модели с пробной массой. Рисунок из статьи ATLAS [2].

На рисунке 13.1 минимум около 126 ГэВ достигает локальной значимости 5.9σ . Вертикальная шкала логарифмическая. Красные линии переводят её в односторонние Z : чем глубже минимум p -значения, тем сильнее несогласие с фоном.

Пунктирная кривая отвечает другому расчёту: какой результат ожидается, если сигнал Стандартной модели действительно есть при каждой проверяемой массе. Она не является подгонкой чёрной кривой.

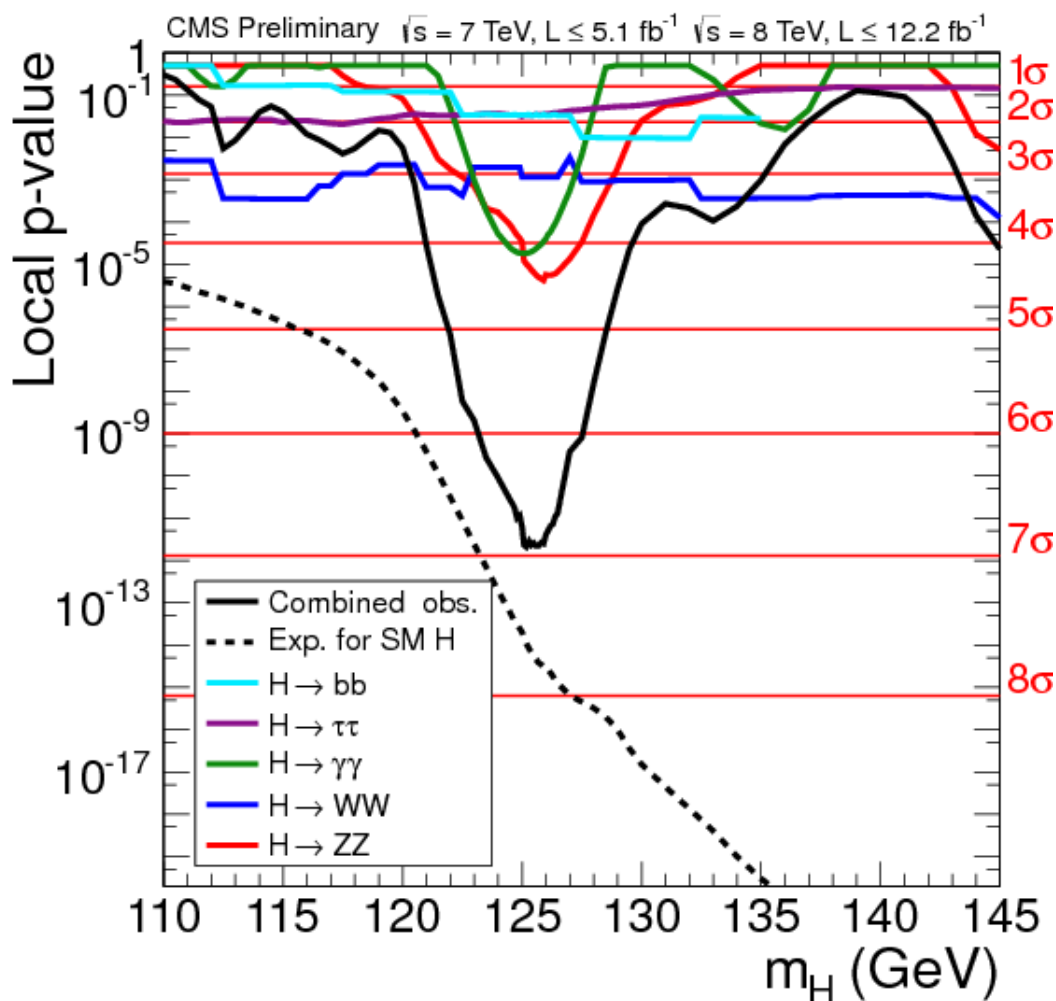


Рис. 13.2. Локальные p -значения в предварительном результате CMS с данными при 7 и 8 ТэВ и светимостью до 12.2 fb^{-1} при 8 ТэВ. Цветные кривые показывают отдельные каналы распада, чёрная — их объединение. Источник: CMS PAS HIG-12-045 [3].

В первой статье об открытии CMS сообщала локальную значимость 5.0σ [4]. На рисунке 13.2 используется уже больший набор данных, и минимум стал глубже. Сравнивая результаты, нужно проверять накопленную светимость и версию анализа: число сигма относится к конкретной выборке.

Обе иллюстрации относятся к проверке **фоновой гипотезы**. Качество согласия модели «сигнал плюс фон» с данными проверяется отдельно. Малое p -значение фона само по себе не сообщает, насколько подробно выбранная модель сигнала описывает все свойства событий.

13.8.1. Локальный результат и поиск по массе

При фиксированной заранее массе получается локальное p -значение. Если массу выбирали по самому глубокому минимуму в диапазоне, в повторяемом опыте нужно каждый раз просматривать весь этот диапазон. Так определяется глобальное p -значение: вероятность найти где-либо в нём столь же сильное отклонение [5].

Соседние массы проверяются по одним и тем же данным, поэтому их результаты коррелированы. Число точек на нарисованной кривой не равно числу независимых попыток.

13.8.2. Откуда возникает правило $Z = \sqrt{q_0}$

Покажем его на простой гауссовой модели. Пусть измеряется амплитуда сигнала $Y \sim \mathcal{N}(s, \sigma^2)$, где σ известна, а физически $s \geq 0$. Проверяется фон $s = 0$. Подгонка с ограничением даёт $\hat{s} = \max(0, y)$. Улучшение суммы квадратов равно

$$q_0 = \chi^2(0) - \chi^2(\hat{s}) = \begin{cases} y^2/\sigma^2, & y > 0, \\ 0, & y \leq 0. \end{cases} \quad (13.26)$$

При верном фоне половина опытов имеет $Y \leq 0$ и даёт $q_0 = 0$. Положительная часть соответствует половине распределения χ_1^2 . Для полученного $q_0 > 0$ имеем

$$p_0 = P_{s=0}(q_0(Y) \geq q_0) = 1 - \Phi(\sqrt{q_0}), \quad Z = \sqrt{q_0}. \quad (13.27)$$

Поэтому $q_0 = 25$ даёт односторонние 5σ . Для чистого χ_1^2 вероятность правее 25 вдвое больше: квадрат учитывает оба знака исходного гауссова отклонения.

В задачах поиска частиц аналогичная формула возникает как асимптотика профильной статистики при одном неотрицательном параметре сигнала и выполнении соответствующих условий регулярности [5]. К произвольному минимуму χ^2 она не относится.

13.9. Когда распределение нужно получить моделированием

13.9.1. Малые пуассоновские счёты

Рассмотрим пример из лекции: двадцать независимых пуассоновских бинов с заранее заданными ожиданиями μ_i . Здесь нет подгоняемых параметров и не фиксируется суммарное число событий:

$$N_i \sim \text{Pois}(\mu_i), \quad T_P = \sum_{i=1}^{20} \frac{(N_i - \mu_i)^2}{\mu_i}. \quad (13.28)$$

Это **статистика Пирсона**. При больших ожиданиях каждое $(N_i - \mu_i)/\sqrt{\mu_i}$ приближается к стандартной гауссовой величине, и сумма квадратов приближённо имеет распределение χ_{20}^2 . Но в примере ожидания малы:

Таблица 13.4. Ожидаемые числа событий в двадцати бинах

Бины	μ_i	Бины	μ_i
1, 20	0.158	6, 15	1.785
2, 19	0.295	7, 14	2.355
3, 18	0.513	8, 13	2.899
4, 17	0.833	9, 12	3.330
5, 16	1.263	10, 11	3.568

У статистики 13.28 среднее всё равно равно 20. Для каждого бина $E[(N_i - \mu_i)^2] = \mu_i$. Но совпадения среднего недостаточно, чтобы совпало распределение. Используя центральный четвёртый момент Пуассона $E[(N_i - \mu_i)^4] = \mu_i + 3\mu_i^2$, получаем

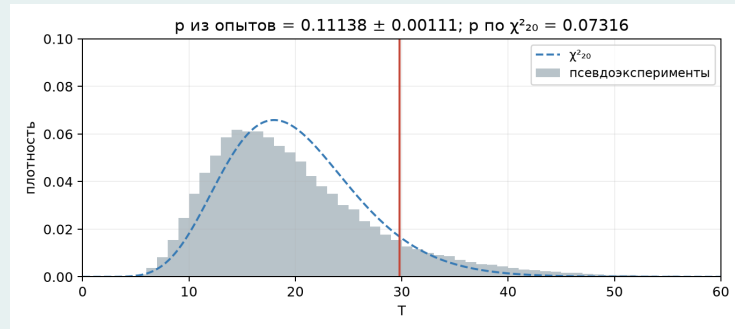
$$\text{Var}(T_P) = \sum_{i=1}^{20} \left(2 + \frac{1}{\mu_i} \right) \approx 71.14. \quad (13.29)$$

У χ^2_{20} дисперсия равна 40. Уже по ширине видно, что аналитическая замена неточна. Малые μ_i дают особенно большой добавочный вклад.

13.9.2. Интерактив: распределение из опытов

Для каждого опыта генерируем двадцать счётов из заданных пуассоновских распределений и вычисляем ту же сумму Пирсона. Полученные значения задают распределение T_P при нашей модели.

Ползунок задаёт проверяемое значение t_{obs} . При $t_{\text{obs}} = 29.8$ независимая проверка на миллионе опытов даёт $p \approx 0.113$ со статистической ошибкой моделирования около 0.0003. Формула для χ^2_{20} даёт 0.07316. Апплет использует 80 000 опытов, поэтому последняя цифра его результата может отличаться.



Интерактив. Проверка χ^2 -приближения при малых счётах

Гистограмма построена по 80 000 опытов для ожиданий из таблицы 13.4. Пунктир — плотность χ^2_{20} , вертикальная линия — $t_{\text{obs}} = 29.8$. Оба p -значения вычисляются по всему хвосту, включая значения за правой границей рисунка.



Можно выбрать другую статистику, например отношение пуассоновских правдоподобий к насыщенной модели. В насыщенной модели для каждого бина разрешено отдельное среднее, и его оценка равна n_i :

$$T_L = 2 \sum_i \left[\mu_i - n_i + n_i \ln \frac{n_i}{\mu_i} \right]. \quad (13.30)$$

При $n_i = 0$ логарифмическое слагаемое по пределу равно нулю. Если n_i близко к μ_i , разложение логарифма даёт статистику Пирсона в ведущем квадратичном порядке. При малых счётах обе статистики определены, но имеют разные распределения. Само использование правдоподобия не делает распределение 13.30 точно равным χ^2 при любом числе событий.

13.9.3. Что повторять в псевдоэксперименте

Если параметры подгонялись по данным, то и в каждом псевдоэксперименте их нужно подгонять заново. Если в анализ входили вспомогательные калибровки, отбор событий или поиск по диапазону масс, процедура калибровки статистики должна учитывать эти действия.

При полностью заданной гипотезе H_0 доля превышений оценивается как

$$\hat{p} = \frac{k}{B}, \quad \sigma_{\hat{p}} \approx \sqrt{\frac{\hat{p}(1 - \hat{p})}{B}}, \quad (13.31)$$

где B — число независимых опытов, k — число значений $T \geq t_{\text{obs}}$. Формула

ошибки годится, когда превышений и непревышений достаточно много. Если $k = 0$, нулевая оценка ошибки по этой формуле бессодержательна: редкий хвост просто не разрешён данной выборкой. При $p \approx 2.9 \cdot 10^{-7}$ даже миллион опытов даёт в среднем меньше одного превышения.

Для численного рангового теста с конечной калибровочной выборкой часто используют $(k + 1)/(B + 1)$, включая сам проверяемый результат в набор из $B + 1$ равноценных реализаций. Именно так устроен апплет распределения p -значений выше. Этот приём не заменяет проверку модели и не устраняет неопределённость редкого хвоста.

Если H_0 содержит неизвестные параметры, остаётся вопрос, при каких значениях генерировать опыты. Подстановка найденных оценок даёт распространённую приближённую калибровку. Её частотные свойства нужно проверять по допустимым значениям мешающих параметров. Одна генерация при наилучшем фите не гарантирует правильный уровень для всей составной гипотезы.

Итоги главы

- В линейной гауссовой задаче с известной ковариацией минимум χ^2 имеет $N - m$ степеней свободы. Корреляции учитываются ковариацией; подгонка уменьшает размерность пространства остатков.
- p -значение — хвостовая вероятность выбранной статистики при проверяемой гипотезе. Отношение χ^2/ν без самого ν не определяет эту вероятность.
- «Число сигма» задаёт гауссово представление p -значения. Один и два хвоста, проверка фона и качество фита требуют разных постановок; их нельзя смешивать.
- При неточной аналитической калибровке распределение статистики получают псевдоэкспериментами с той же процедурой анализа. Число опытов ограничивает точность вычисления редких хвостов.

13.10. Задачи

Задача 1. Две статистики для пуассоновских бинов

В двадцати независимых бинах $N_i \sim \text{Pois}(\mu_i)$. Первые десять ожиданий равны (0.158, 0.295, 0.513, 0.833, 1.263, 1.785, 2.355, 2.899, 3.330, 3.568), остальные заданы симметрией $\mu_{21-i} = \mu_i$. Все ожидания известны; параметры по данным не подгоняются.

1. Запишите статистику Пирсона и выведите статистику отношения правдоподобий к насыщенной модели, где каждому бину разрешено своё среднее.
2. Разложите логарифм при $n_i = \mu_i + \delta_i$, $|\delta_i| \ll \mu_i$. Сравните ведущие квадратичные члены.
3. Найдите вклады пустого бина в обе статистики. Почему малое число событий само по себе не мешает вычислить эти вклады?
4. Выведите среднее и дисперсию статистики Пирсона, используя $E[(N_i - \mu_i)^4] = \mu_i + 3\mu_i^2$.
5. Сгенерируйте 50 000 опытов. Сравните распределения обеих статистик с χ^2_{20} . Повторите при увеличении всех μ_i в 100 раз.

Задача 2. Степени свободы до и после подгонки

Даны $N = 12$ независимых измерений $Y_i \sim \mathcal{N}(\mu, \sigma^2)$ при $\mu = 10$ каналах и известной $\sigma = 2$ канала. Проведите 20 000 псевдоэкспериментов.

1. В каждом опыте вычислите среднее $\bar{Y}$, сумму квадратов относительно истинного μ и минимальную сумму квадратов относительно $\bar{Y}$; ошибки в знаменателях равны σ .
2. Проверьте точное разложение суммы квадратов на два слагаемых: минимум и $N(\bar{Y} - \mu)^2/\sigma^2$.
3. Сравните три распределения с χ_{12}^2 , χ_{11}^2 и χ_1^2 . Проверьте их средние и дисперсии.
4. Объясните, почему остатки после подгонки не являются двенадцатью независимыми стандартными гауссовыми величинами.
5. Найдите р-значение для $\chi_{\min}^2 = 12$. Что получится, если ошибочно оставить двенадцать степеней свободы?

Задача 3. Одинаковый χ^2 и разная форма остатков

Построим два учебных набора калибровочных данных. Энергии $E_i = i$ МэВ, $i = 0, \dots, 11$, ошибки амплитуды — по 5 каналов. Матрица X содержит столбцы 1 и E_i .

Для первого набора сгенерируйте двенадцать независимых чисел $w_i \sim \mathcal{N}(0, 1)$ и удалите их проекцию на прямую: $\mathbf{v} = \mathbf{w} - X(X^T X)^{-1} X^T \mathbf{w}$. Для второго положите $v_i = (E_i - 5.5)^2 - \frac{1}{12} \sum_j (E_j - 5.5)^2$, используя числовые значения энергий в МэВ.

В каждом случае нормируйте остаток: $\mathbf{u} = \sqrt{12} \mathbf{v} / \|\mathbf{v}\|$, и задайте амплитуды $y_i = 2 + 100E_i + 5u_i$ в каналах.

1. Подгоните оба набора прямой с двумя свободными параметрами.
2. Проверьте, что в обоих случаях $\chi_{\min}^2 = 12$ и р-значения совпадают.
3. Постройте нормированные остатки по энергии. Какую особенность второго набора скрывает одно число χ^2 ?
4. Объясните, почему эти специально сконструированные наборы нельзя использовать как случайную выборку для проверки распределения $\chi_{\min}^2$.

Задача 4. Когда р-значения равномерны

Сгенерируйте 20 000 независимых значений $T \sim \chi_{10}^2$.

1. Вычислите для каждого $p = P(\chi_{10}^2 \geq T)$, постройте гистограмму и найдите долю $p \leq 0.05$.
2. Повторите расчёт, заменив каждое T на $T/0.8^2$, но сохранив калибровочное распределение χ_{10}^2 . Какой ошибке в задании стандартных отклонений соответствует такая замена?
3. Сгенерируйте отдельную калибровочную выборку из 20 000 значений χ_{10}^2 . Для исходных T вычислите ранговые р-значения $(k+1)/(B+1)$, где k — число калибровочных значений не меньше T , $B = 20000$.
4. Повторите опыт для $N \sim \text{Pois}(2)$, используя точное $p(n) = P(N \geq n)$. Найдите наименьшее целое n , при котором $p(n) \leq 0.05$, и точную вероятность отклонения верной гипотезы.
5. Почему в опыте с Пуассоном доля отклонений не обязана быть ровно 5% даже при неограниченном числе псевдоэкспериментов?

Задача 5. Один хвост, два хвоста и открытие

1. Вычислите $1 - \Phi(Z)$ и $2[1 - \Phi(Z)]$ для $Z = 1, \dots, 5$. Проверьте таблицы главы.
2. Найдите оба перевода в Z для $p = 0.05$, $2.7 \cdot 10^{-3}$ и $2.87 \cdot 10^{-7}$.
3. Для $T = 1200$ при $T \sim \chi^2_{1000}$ сравните точный односторонний перевод p -значения с $(T - 1000)/\sqrt{2000}$.
4. В модели $Y \sim \mathcal{N}(s, \sigma^2)$ при известной σ , $s \geq 0$, выведите $q_0 = \max(0, Y/\sigma)^2$ для проверки $s = 0$. Найдите вероятность $q_0 \geq 25$ при фоне. Сравните с $P(\chi^2_1 \geq 25)$.
5. Исследователь решил считать отклонение односторонним, но сторону выбирает по знаку результата. Если в каждой выбранной стороне он использует порог 5%, какова полная вероятность ложного отклонения верной гауссовой гипотезы?

Литература

- [1] Cowan, G. *Statistical Data Analysis*. Oxford University Press. 1998.
- [2] ATLAS Collaboration. Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC. *Physics Letters B*, **716**, 1–29. 2012. doi:10.1016/j.physletb.2012.08.020.
- [3] CMS Collaboration. Combination of Standard Model Higgs Boson Searches and Measurements of the Properties of the New Boson with a Mass near 125 GeV. 2012. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/Hig12045TWiki>.
- [4] CMS Collaboration. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Physics Letters B*, **716**, 30–61. 2012. doi:10.1016/j.physletb.2012.08.021.
- [5] Cowan, G., Cranmer, K., Gross, E., и Vitells, O. Asymptotic Formulae for Likelihood-Based Tests of New Physics. *The European Physical Journal C*, **71**, 1554. 2011. doi:10.1140/epjc/s10052-011-1554-0. <https://arxiv.org/abs/1007.1727>.