

χ^2 -интервалы и систематические неопределённости

Содержание

12.1.	Физическая задача: два измерения с общей нормировкой	161
12.2.	Общий случай с ковариационной матрицей	161
12.3.	Интервалы из формы χ^2	163
12.4.	Откуда берутся систематические неопределённости	165
12.5.	Профилирование мешающих параметров	166
12.6.	Возвращаемся к общей нормировке	167
12.7.	Парадокс Д'Агостини	169
12.8.	Распространение ошибок после подгонки	170
12.9.	Что проверить перед записью результата	171
12.10.	Задачи	172
	Литература	173

Аннотация

В прошлой главе ошибки измерений были независимы. Теперь включим в подгонку корреляции и неизвестные калибровочные параметры. Разберём, как из формы χ^2 получают ошибки и интервалы, откуда берутся штрафные члены и когда профилирование можно выполнить аналитически. Пример с общей нормировкой покажет, почему при построении ковариации нужно различать данные и модель.

12.1. Физическая задача: два измерения с общей нормировкой

Пусть сечение процесса измерено по двум независимым наборам событий. Оба результата используют одну калибровку светимости, поэтому часть их ошибки общая. Возьмём числа из примера Д’Агостини [1] и для определённости будем измерять сечение в пб:

$$y_1 = (8.00 \pm 0.16) \text{ пб}, \quad y_2 = (8.50 \pm 0.17) \text{ пб}. \quad (12.1)$$

Выписанные ошибки относятся к независимой части. Кроме них имеется общая относительная неопределённость нормировки $f_N = 0.10$. В расчёте считаем независимые ошибки гауссовыми, а их стандартные отклонения $s_1 = 0.16$ пб и $s_2 = 0.17$ пб — заданными числами. Они не меняются вместе с пробным сечением. Числа относятся к учебной модели.

Как объединить эти два числа? Если пока забыть про нормировку, взвешенное среднее равно 8.235 пб. После учёта общей нормировки наша модель даст

$$\hat{\mu} = (8.235 \pm 0.832) \text{ пб}, \quad (12.2)$$

где ошибка найдена по кривизне в минимуме. Но есть весьма правдоподобный способ записать ковариационную матрицу, который вместо этого приводит к 7.87 ± 0.81 пб. Центральное значение оказалось ниже обеих точек.

Почему так получилось? И какое усреднение описывает наш эксперимент? Для ответа нужно проследить, как общая калибровка входит в предсказание. Начнём с МНК для коррелированных измерений.

12.2. Общий случай с ковариационной матрицей

Соберём N результатов в вектор $\mathbf{y}$, а их математические ожидания — в вектор $\lambda(\theta)$. Параметры модели обозначены θ . Пусть данные имеют многомерное гауссово распределение с известной положительно определённой ковариацией V_y :

$$\mathbf{Y} \sim \mathcal{N}(\lambda(\theta), V_y). \quad (12.3)$$

При фиксированной V_y логарифм правдоподобия с точностью до постоянного слагаемого определяется квадратичной формой [2]

$$\chi^2(\theta) = (\mathbf{y} - \lambda)^T V_y^{-1} (\mathbf{y} - \lambda). \quad (12.4)$$

В главе о двумерном гауссе мы уже встречали такую меру отклонения от центра. Теперь центром служит модель, которую мы подгоняем. Если V_y диагональна, формула 12.4 превращается в знакомую сумму $\sum_i r_i^2 / \sigma_i^2$.

При корреляциях отдельные остатки нельзя рассматривать независимо. Например, общая ошибка шкалы может сдвинуть сразу несколько точек в одну сторону. Вероятность такого согласованного отклонения отличается от вероятности нескольких независимых сдвигов того же размера. Внедиагональные элементы V_y как раз учитывают эту связь.

Если ковариация самой модели данных зависит от параметров, из нормировки гауссовой плотности появляется ещё одно слагаемое:

$$-2 \ln L(\theta) = \mathbf{r}^T V_y(\theta)^{-1} \mathbf{r} + \ln \det V_y(\theta) + \text{const}, \quad \mathbf{r} = \mathbf{y} - \lambda(\theta). \quad (12.5)$$

Отбрасывать меняющийся определитель нельзя. Ниже появится внешне похожая матрица после профилирования. Мы отдельно выясним, почему к ней это правило автоматически не переносится.

12.2.1. Несколько линейных параметров

Пусть модель линейна по p параметрам:

$$\lambda = X\theta, \quad X \in \mathbb{R}^{N \times p}. \quad (12.6)$$

Столбцы известной матрицы X задают, как каждый параметр меняет предсказания. Для прямой это столбцы 1 и x_i . Приравнявая производные χ^2 к нулю, получаем нормальные уравнения

$$X^T V_y^{-1} X \hat{\theta} = X^T V_y^{-1} \mathbf{y}. \quad (12.7)$$

Если столбцы X линейно независимы, решение единственно:

$$\hat{\theta} = (X^T V_y^{-1} X)^{-1} X^T V_y^{-1} \mathbf{y}, \quad V_\theta = (X^T V_y^{-1} X)^{-1}. \quad (12.8)$$

Проверка ковариации повторяет вывод для прямой. Оценка имеет вид $\hat{\theta} = B\mathbf{y}$, где $BX = I$. Поэтому $E[\hat{\theta}] = \theta$ и $\text{Cov}(\hat{\theta}) = BV_y B^T = V_\theta$. Для этих двух равенств достаточно верных первых двух моментов данных. Гауссова модель дополнительно даёт гауссово распределение оценок.

12.2.2. Ошибка по кривизне

В нелинейной модели точного решения обычно нет. Найдём минимум численно и разложим функцию около него. Обозначим через K половину матрицы вторых производных:

$$\chi^2(\theta) \simeq \chi_{\min}^2 + (\theta - \hat{\theta})^T K (\theta - \hat{\theta}),$$

$$K_{ij} = \frac{1}{2} \left. \frac{\partial^2 \chi^2}{\partial \theta_i \partial \theta_j} \right|_{\hat{\theta}}, \quad V_\theta \simeq K^{-1}. \quad (12.9)$$

Множитель $1/2$ приходит из формулы Тейлора. Его можно проверить на одной параболе: вторая производная $(\theta - \hat{\theta})^2 / \sigma^2$ равна $2/\sigma^2$. Если обратить её без двойки, дисперсия получится вдвое меньше.

Для линейной модели с известной ковариацией формула точна. В общем случае K^{-1} даёт локальную оценку ковариации. Она требует изолированного внутреннего минимума с положительной кривизной и достаточно близкой к квадратичной формы правдоподобия в интересующей области. Узкий минимум ещё нужно найти среди всех остальных минимумов, если они есть.

12.3. Интервалы из формы χ^2

12.3.1. Один параметр

Для одного параметра квадратичная форма записывается совсем просто:

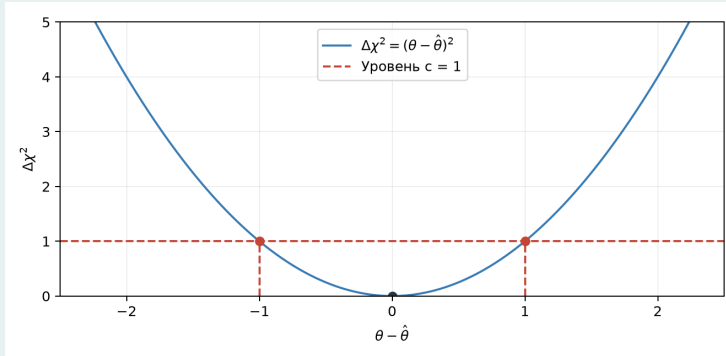
$$\Delta\chi^2(\theta) = \chi^2(\theta) - \chi_{\min}^2 \approx \frac{(\theta - \hat{\theta})^2}{\sigma_{\hat{\theta}}^2}. \quad (12.10)$$

Отступим от минимума на одну стандартную ошибку. Значение χ^2 вырастет на единицу:

$$\Delta\chi^2(\hat{\theta} \pm \sigma_{\hat{\theta}}) = 1. \quad (12.11)$$

Если вместе с θ подгоняются другие неизвестные параметры, здесь нужна **профильная** функция: при каждом пробном θ остальные параметры подбираются заново. Сечение с зафиксированными параметрами может оказаться значительно уже. Мы видели это на коррелированной паре в главе 10.

В апплете меняются стандартная ошибка $\sigma_{\hat{\theta}}$ и уровень отсечения c . Обе координаты безразмерные. Пересечения удовлетворяют $\theta - \hat{\theta} = \pm\sqrt{c}\sigma_{\hat{\theta}}$. Попробуйте поднять уровень с 1 до 4: интервал расширится вдвое.



Интерактив. Интервал по пересечению с уровнем χ^2

При $\sigma_{\hat{\theta}} = 1$ и $c = 1$ границы лежат при $\theta - \hat{\theta} = \pm 1$. Пунктир задаёт уровень отсечения, сплошная кривая показывает $\Delta\chi^2$.



В линейной гауссовой задаче с известными дисперсиями интервал 12.11 имеет точное покрытие около 68.3%. Для нелинейной модели или негауссовых данных правило $\Delta\chi^2 = 1$ само по себе такого покрытия не гарантирует. Одинаковая высота над минимумом может соответствовать разным вероятностям в ансамбле экспериментов.

12.3.2. Совместная область нескольких параметров

Для p параметров условие

$$(\theta - \hat{\theta})^T V_{\hat{\theta}}^{-1} (\theta - \hat{\theta}) \leq c \quad (12.12)$$

задаёт эллипсоид. Его оси определяются собственными векторами ковариации, а размеры — её собственными значениями и выбранным c . На рисунке 12.1 показаны два сечения квадратичной поверхности постоянным уровнем.

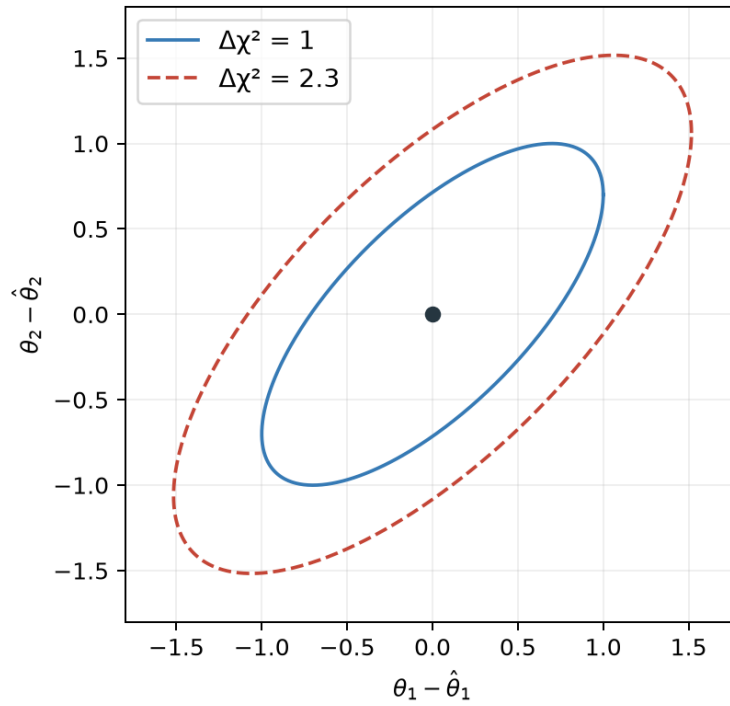


Рис. 12.1. Контуры квадратичной формы для двух параметров с единичными стандартными ошибками и корреляцией $\rho = 0.7$. Сплошной контур соответствует $s = 1$, пунктирный — $s = 2.30$. Отложены отклонения параметров от минимума.

Порог для совместной области зависит от числа параметров. Для двух гауссовых оценок область с $s = 1$ покрывает истинную пару всего примерно в 39.3% опытов. Для покрытия 68.3% требуется $s \approx 2.30$. Это тот же результат, который мы получили для двумерного гаусса. Число измерений N здесь не заменяет число параметров p .

12.3.3. Когда эллипса недостаточно

В осцилляционной задаче разные частоты могут давать близкие предсказания в точках измерения. Для иллюстрации возьмём четыре безразмерных момента $t_i = i, i = 1, \dots, 4$, и значения $y_i = 0.6 \sin(1.2t_i)$. Это специально заданные точки без добавленной случайной флуктуации. Приписав им независимые гауссовы ошибки $s = 0.15$, получим функцию

$$\chi^2(A, \omega) = \sum_{i=1}^4 \frac{[y_i - A \sin(\omega t_i)]^2}{s^2}. \quad (12.13)$$

На рисунке 12.2 видны две разделённые области. У этой учебной модели есть точная неоднозначность: при целых t_i частоты ω и $\omega + 2\pi$ дают те же значения синуса.

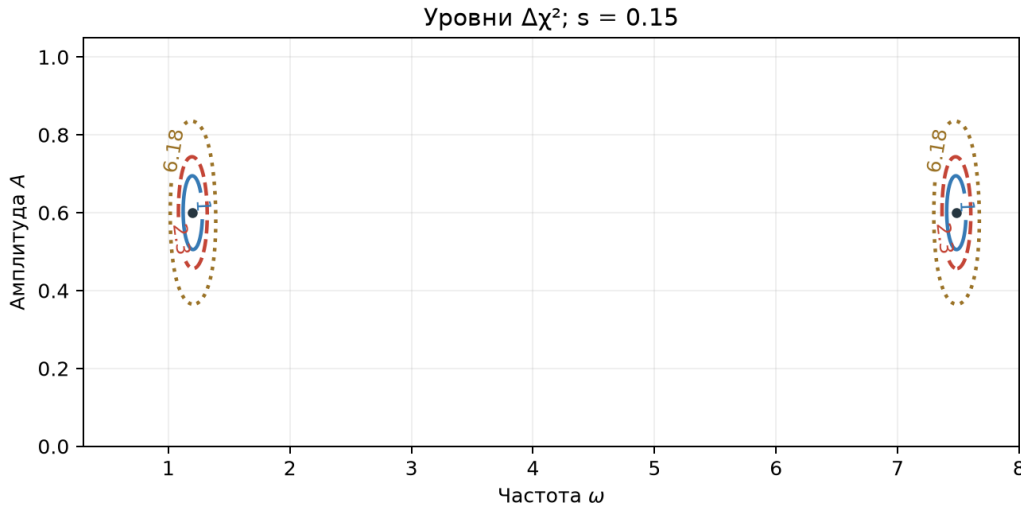


Рис. 12.2. Контурсы $\Delta\chi^2 = 1, 2.30$ и 6.18 для четырёх точек осцилляционной модели. Минимумы при $A = 0.6$, $\omega = 1.2$ и $\omega = 1.2 + 2\pi$ дают одинаковые предсказания. Локальный эллипс около одного минимума не описывает вторую область.

Можно очень точно вычислить вторые производные около первого минимума и полностью пропустить второй. Поэтому в нелинейной задаче следует исследовать весь допустимый диапазон параметров. Сами нарисованные контуры ещё не имеют гарантированного доверительного уровня: его проверяют по распределению статистики. Этим займёмся в следующей главе.

12.4. Откуда берутся систематические неопределённости

Вернёмся к эксперименту. Сечение зависит от светимости и эффективности, энергия — от калибровки детектора, число сигнальных событий — от описания фона. Все эти величины входят в модель. Если они известны неточно, их неопределённость передаётся результату.

Параметры, ради которых проводится измерение, называются **параметрами интереса**. Остальные неизвестные параметры, влияющие на предсказание, — **мешающие параметры**. Например, в измерении сечения светимость является мешающим параметром. В отдельной работе по калибровке светимости она сама станет параметром интереса. Название определяется задачей анализа.

Обозначим эти две группы через θ и η :

$$\lambda = \lambda(\theta, \eta). \quad (12.14)$$

Систематикой в разговоре называют источник неопределённости: шкалу энергии, нормировку потока, форму фона. Для расчёта нужно уточнить, как именно этот источник меняет λ и какими данными ограничены его параметры. Одного числа «систематическая ошибка 5%» для произвольной подгонки мало.

12.4.1. Вспомогательное измерение и штраф

Пусть независимая калибровка дала вектор $\mathbf{z}$. В гауссовом приближении её вероятностная модель имеет вид

$$\mathbf{Z} \sim \mathcal{N}(\eta, V_\eta), \quad L(\theta, \eta) = L_{\text{main}}(\mathbf{y} \mid \theta, \eta) L_{\text{aux}}(\mathbf{z} \mid \eta). \quad (12.15)$$

Матрица V_η здесь описывает разброс **результата калибровки** $\mathbf{Z}$ при фиксированном η . В частотной постановке сам неизвестный параметр не

становится случайным. Запись « $\eta = \eta_0 \pm s_\eta$ » означает результат вспомогательного измерения; его полученное значение η_0 играет роль z .

При известных постоянных ковариациях и независимости двух наборов данных произведение правдоподобий даёт сумму

$$\chi^2(\theta, \eta) = [\mathbf{y} - \lambda(\theta, \eta)]^T V_y^{-1} [\mathbf{y} - \lambda(\theta, \eta)] + (\eta - \mathbf{z})^T V_\eta^{-1} (\eta - \mathbf{z}). \quad (12.16)$$

Второе слагаемое называют **штрафным членом**. Сдвиг одного калибровочного параметра на его стандартную ошибку добавляет единицу к χ^2 , если других коррелированных калибровочных параметров нет. Основные данные могут предпочесть такой сдвиг. Совместная подгонка сравнит улучшение описания основных данных с ухудшением описания калибровки.

Если внешней информации нет, соответствующий штраф отсутствует: параметр определяется только основными данными. Данных может оказаться недостаточно, например для разделения общей нормировки и самого сечения. Если калибровка использует те же события, что и основной анализ, перемножать правдоподобия как независимые нельзя — общие данные будут учтены дважды.

Не всякая систематика получена из гауссовой калибровки. Интервал между двумя теоретическими моделями сам по себе не задаёт ни гауссову ошибку, ни стандартное отклонение. Для такого ограничения требуется отдельное обоснование; формула 12.16 его не создаёт.

12.5. Профилирование мешающих параметров

При каждом фиксированном θ найдём минимум по η :

$$\chi_p^2(\theta) = \min_{\eta} \chi^2(\theta, \eta) = \chi^2(\theta, \hat{\eta}(\theta)). \quad (12.17)$$

Двойная шляпка обозначает результат этой условной подгонки. При изменении параметра интереса калибровочные параметры могут сдвигаться, частично компенсируя изменение предсказания. Поэтому профиль обычно шире сечения, в котором все η удерживаются в точке общего минимума.

В квадратичном приближении тот же эффект виден из матрицы K . Разобьём её на блоки по двум группам параметров:

$$K = \begin{pmatrix} K_{\theta\theta} & K_{\theta\eta} \\ K_{\eta\theta} & K_{\eta\eta} \end{pmatrix}, \quad K_p = K_{\theta\theta} - K_{\theta\eta} K_{\eta\eta}^{-1} K_{\eta\theta}. \quad (12.18)$$

Подстановка минимума по η даёт кривизну профиля K_p . Обратная матрица K_p^{-1} совпадает с $\theta\theta$ -блоком полной матрицы K^{-1} . Обратить только блок $K_{\theta\theta}$ — значит получить ошибку при фиксированных мешающих параметрах. Это другой расчёт.

12.5.1. Линейная зависимость от систематики

Часть минимизации можно выполнить аналитически. При фиксированном θ разложим предсказание около результата калибровки $\mathbf{z}$:

$$\lambda(\theta, \eta) \simeq \lambda_0(\theta) + H(\theta)\delta, \quad \delta = \eta - \mathbf{z}, \quad H_{ij} = \left. \frac{\partial \lambda_i}{\partial \eta_j} \right|_{\eta=\mathbf{z}}. \quad (12.19)$$

Здесь $\lambda_0 = \lambda(\theta, \mathbf{z})$, а H содержит N строк и столько столбцов, сколько имеется

мешающих параметров. Каждый столбец показывает согласованное изменение всех предсказаний при сдвиге одного параметра.

Вывод

12.5.2. Аналитический минимум

Положим $W = V_y^{-1}$, $C = V_\eta$ и $\mathbf{r} = \mathbf{y} - \lambda_0$. Тогда

$$\chi^2 = (\mathbf{r} - H\delta)^T W (\mathbf{r} - H\delta) + \delta^T C^{-1} \delta. \quad (12.20)$$

Производная по δ даёт систему

$$(C^{-1} + H^T W H) \hat{\delta} = H^T W \mathbf{r}. \quad (12.21)$$

Обозначим матрицу слева через M . В минимуме $\hat{\delta} = M^{-1} H^T W \mathbf{r}$. Подставим это решение в исходную сумму квадратов:

$$\chi_p^2 = \mathbf{r}^T (W - W H M^{-1} H^T W) \mathbf{r}. \quad (12.22)$$

Матрицу в скобках можно записать как $(V_y + H C H^T)^{-1}$. Равенство проверяется умножением на $V_y + H C H^T$. Получаем

$$\chi_p^2(\theta) = \mathbf{r}^T V^{-1} \mathbf{r}, \quad V = V_y + H V_\eta H^T. \quad (12.23)$$

Второе слагаемое в V знакомо по распространению ошибок. Если одна калибровка влияет на много точек, её вклад обычно недиагонален. Добавить соответствующие дисперсии только к диагонали — значит потерять эту связь.

Для точно линейной зависимости от η и гауссова штрафа формула 12.23 точна при каждом фиксированном θ . Если использовано лишь линейное разложение, нужно проверить его во всём диапазоне сдвигов, допускаемых подгонкой.

12.5.3. Нужно ли теперь добавлять $\ln \det V$?

Нет. В формуле 12.23 мы уже выполнили минимизацию исходного χ^2 . Его нормировочные множители определялись постоянными V_y и V_η . Определитель эффективной матрицы после минимизации не появился, даже если H зависит от θ .

У формулы 12.5 другая исходная постановка: там зависимость ковариации от параметров задана в самой плотности данных. Интегрирование гауссова выражения по мешающим параметрам также создаёт зависящий от ширины множитель. Профилирование такого интегрирования не выполняет. Чтобы выбрать нужное выражение, следует вернуться к исходному правдоподобию.

12.6. Возвращаемся к общей нормировке

Запишем модель для данных 12.1:

$$Y_i = (1 + \alpha)\mu + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, s_i^2), \quad Z \sim \mathcal{N}(\alpha, f_N^2). \quad (12.24)$$

Все ε_i и результат калибровки Z независимы. Полученное значение калибровки равно $z = 0$, то есть номинальный масштаб равен единице. Параметр α безразмерный и один для обеих точек. Линейная запись масштаба соответствует нашей модели относительной неопределённости; интересующая область находится далеко от нефизического значения $1 + \alpha \leq 0$.

Полная функция для подгонки равна

$$\chi^2(\mu, \alpha) = \sum_i \frac{[y_i - (1 + \alpha)\mu]^2}{s_i^2} + \frac{\alpha^2}{f_N^2}. \quad (12.25)$$

Введём статистически взвешенное среднее m , его дисперсию v при фиксированной нормировке и сумму квадратов отклонений от этого среднего:

$$S = \sum_i \frac{1}{s_i^2}, \quad m = \frac{1}{S} \sum_i \frac{y_i}{s_i^2}, \quad v = \frac{1}{S}, \quad q_0 = \sum_i \frac{(y_i - m)^2}{s_i^2}. \quad (12.26)$$

Поскольку $\sum_i (y_i - m)/s_i^2 = 0$, функция распадается на три слагаемых:

$$\chi^2(\mu, \alpha) = q_0 + \frac{[m - (1 + \alpha)\mu]^2}{v} + \frac{\alpha^2}{f_N^2}. \quad (12.27)$$

Последние два слагаемых неотрицательны и одновременно обращаются в нуль при $\hat{\alpha} = 0$, $\hat{\mu} = m$. Это и есть общий минимум. Две точки определяют произведение $(1 + \alpha)\mu$, а разделить его на масштаб и сечение позволяет вспомогательная калибровка.

12.6.1. Профиль и ошибка сечения

При фиксированном μ минимум по α имеет вид

$$\hat{\alpha}(\mu) = \frac{f_N^2 \mu (m - \mu)}{v + f_N^2 \mu^2}, \quad \chi_p^2(\mu) = q_0 + \frac{(m - \mu)^2}{v + f_N^2 \mu^2}. \quad (12.28)$$

Профиль уже не парабола: его знаменатель зависит от μ . Вблизи минимума достаточно подставить в знаменатель $\mu = m$. Отсюда локальная ошибка

$$\sigma_\mu^2 \simeq v + f_N^2 m^2. \quad (12.29)$$

Первое слагаемое уменьшается при накоплении статистики. Второе остаётся, пока точность общей калибровки прежняя. Два набора событий не превращают одну калибровку в две независимые.

Для наших чисел получаем

$$m = 8.23486 \text{ пб}, \quad v = 0.013575 \text{ пб}^2, \quad f_N^2 m^2 = 0.678130 \text{ пб}^2. \quad (12.30)$$

Это объясняет результат 12.2. Полная локальная ковариация по (μ, α) равна

$$V_{\mu, \alpha} \simeq \begin{pmatrix} v + f_N^2 m^2 & -f_N^2 m \\ -f_N^2 m & f_N^2 \end{pmatrix}. \quad (12.31)$$

Верхний левый элемент имеет размерность пб², внедиагональный — пб, нижний правый безразмерный. Отрицательная ковариация соответствует модели: увеличение масштаба компенсируется уменьшением сечения.

Для границ по $\Delta\chi_p^2 = 1$ можно использовать весь профиль 12.28. При $f_N < 1$ решение имеет вид

$$\mu_{\pm} = \frac{m \pm \sqrt{f_N^2 m^2 + (1 - f_N^2)v}}{1 - f_N^2}. \quad (12.32)$$

Интервал получится асимметричным. Формула для границ точна для заданной функции, но её покрытие в нелинейной модели требует отдельной проверки. Симметричная ошибка 0.832 пб описывает только местную кривизну.

12.6.2. Та же задача через ковариацию

Производная предсказания по α равна μ для каждой точки. По формуле 12.23

$$V_{ij}(\mu) = s_i^2 \delta_{ij} + f_N^2 \mu^2. \quad (12.33)$$

Весь нормировочный вклад пропорционален матрице из единиц. Подстановка этой матрицы в $\mathbf{r}^T V(\mu)^{-1} \mathbf{r}$ даёт ровно профиль 12.28. Таким способом можно выполнить ту же подгонку без явного параметра α .

Если зафиксировать V при $\mu = m$, получится парабола с тем же минимумом и той же кривизной. Вдали от минимума она отличается от точного профиля. Эквивалентность центрального значения и локальной ошибки не означает совпадения всей функции.

12.7. Парадокс Д'Агостини

Теперь построим нормировочную матрицу из измеренных значений, заменив μ^2 на $y_i y_j$:

$$\begin{aligned} V_{\text{data}} &= \begin{pmatrix} s_1^2 & 0 \\ 0 & s_2^2 \end{pmatrix} + f_N^2 \begin{pmatrix} y_1^2 & y_1 y_2 \\ y_1 y_2 & y_2^2 \end{pmatrix} \\ &= \begin{pmatrix} 0.6656 & 0.6800 \\ 0.6800 & 0.7514 \end{pmatrix} \text{ пб}^2. \end{aligned} \quad (12.34)$$

С этой фиксированной матрицей минимизируем $(\mathbf{y} - \mu \mathbf{1})^T V_{\text{data}}^{-1} (\mathbf{y} - \mu \mathbf{1})$, где $\mathbf{1}$ — вектор из единиц. Решение для произвольной известной фиксированной матрицы V равно

$$\hat{\mu} = \mathbf{w}^T \mathbf{y}, \quad \mathbf{w} = \frac{V^{-1} \mathbf{1}}{\mathbf{1}^T V^{-1} \mathbf{1}}, \quad \sigma_{\mu}^2 = \frac{1}{\mathbf{1}^T V^{-1} \mathbf{1}}. \quad (12.35)$$

При $V = V_{\text{data}}$ веса составляют примерно $w_1 = 1.253$, $w_2 = -0.253$. Отсюда и получается 7.87 ± 0.81 пб. Ошибка после знака $\pm$ здесь формально вычислена по кривизне: матрица построена из тех же случайных данных, поэтому считать её результат истинным стандартным отклонением оценки без проверки нельзя.

Отрицательные веса сами по себе допустимы при объединении коррелированных измерений. Их сумма равна единице, но каждый вес не обязан лежать между нулём и единицей. В нашей задаче ошибка состоит в конкретной подстановке $f_N^2 y_i y_j$: направление изменения всех предсказаний задаётся общим сечением μ , а здесь оно стало зависеть от отдельных флуктуаций точек.

12.7.1. Как появляется сдвиг

Для двух точек обращение матрицы 12.34 даёт простой ответ:

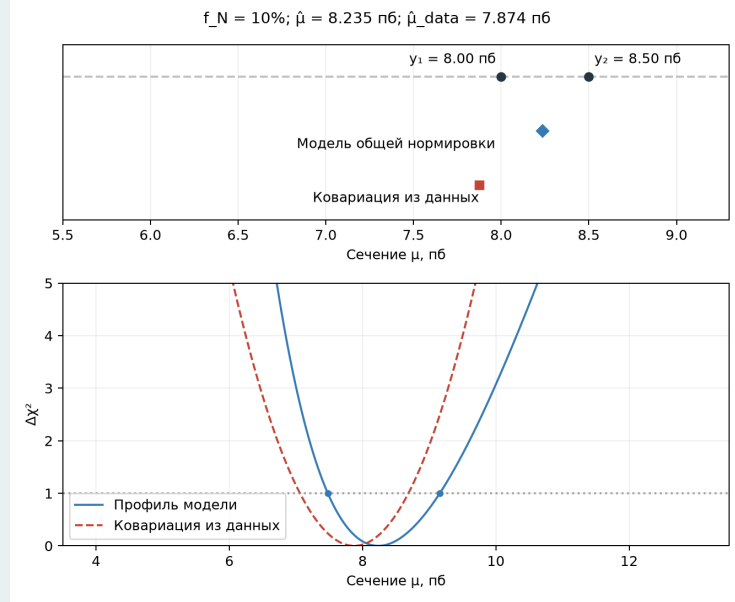
$$\hat{\mu}_{\text{data}} = \frac{m}{1 + f_N^2 q_0}, \quad q_0 = \frac{(y_1 - y_2)^2}{s_1^2 + s_2^2}. \quad (12.36)$$

Чем сильнее расходятся точки, тем больше знаменатель. Для положительного m такая конструкция уменьшает центральное значение. В нашем примере $q_0 \approx 4.59$, и множитель $1/(1 + f_N^2 q_0)$ равен примерно 0.956. Минимизация воспроизводит то, что ей было задано формулой.

Пример разбирается в работе Д'Агостини [1]. Он показывает, почему правила распространения ошибок нельзя переносить в подгонку без проверки зависимости ковариации от данных и параметров.

12.7.2. Интерактив: меняем общую ошибку

Меняйте f_N , оставляя обе точки и их независимые ошибки прежними. На верхнем графике сравниваются центральные значения, полученные двумя способами. На нижнем показаны точный профиль модели с общей нормировкой и квадратичная функция с матрицей из данных. Из каждой функции вычтен её собственный минимум.



Интерактив. Общая нормировка: модель и подстановка данных

При $f_N = 0.10$ модель даёт центральное значение 8.235 пб, а подстановка данных в нормировочную ковариацию — 7.874 пб. Профиль модели асимметричен; пунктирная функция с фиксированной матрицей является параболой.



При $f_N = 0$ обе процедуры совпадают. С ростом f_N профиль расширяется, но его минимум остаётся при t . Минимум функции с V_{data} уходит в сторону меньших значений.

12.8. Распространение ошибок после подгонки

Подогнанные параметры часто нужны для вычисления другой величины $\mathbf{g}(\theta)$. Закон распространения ошибок остаётся тем же, что и в главе 6:

$$\hat{\mathbf{g}} = \mathbf{g}(\hat{\theta}), \quad V_g \simeq J V_{\theta} J^T, \quad J_{ij} = \left. \frac{\partial g_i}{\partial \theta_j} \right|_{\hat{\theta}}. \quad (12.37)$$

Матрица V_{θ} должна включать влияние подогнанных мешающих параметров. Если сама $\mathbf{g}$ зависит ещё и от η , в якобиан и ковариацию включается полный набор (θ, η) . Уже учтённую при подгонке систематику второй раз в квадратуру не добавляют.

Например, после калибровки $y = a + bx$ энергия нового сигнала y_* оценивается как $\hat{E} = (y_* - \hat{a})/\hat{b}$. Если измерение y_* независимо от калибровки и имеет дисперсию s_*^2 , то в линейном приближении

$$\text{Var}(\hat{E}) \simeq \frac{s_*^2 + \text{Var}(\hat{a}) + \hat{E}^2 \text{Var}(\hat{b}) + 2\hat{E} \text{cov}(\hat{a}, \hat{b})}{\hat{b}^2}. \quad (12.38)$$

Ковариационный член может существенно уменьшить ошибку, как и для прямой калибровки в прошлой главе. Если $\hat{b}$ близко к нулю или его ошибка велика, отношение уже нельзя надёжно описывать этим линейным разложением.

12.9. Что проверить перед записью результата

Найденный минимум и его кривизна относятся к той модели, которую мы записали. Если исходные данные — малые пуассоновские счёты, гауссову сумму квадратов нужно заменить соответствующим правдоподобием. Если заметны ошибки по x , модель должна описывать обе измеренные координаты. Например, истинные координаты точек можно ввести как дополнительные мешающие параметры. Одного увеличения вертикальных ошибок для произвольной кривой недостаточно.

При нелинейной подгонке полезно сравнить профиль с параболой, проверить границы допустимых параметров и несколько начальных точек минимизации. Кривизна в одном минимуме не сообщает о других решениях.

В публикации вместе со значениями и единицами параметров нужны их ковариации, использованная функция подгонки и описание внешних ограничений. По этим сведениям другой исследователь сможет повторить расчёт или объединить результат со своим. Для нашей пары чисел особенно существенно, что ошибка нормировки общая: два независимых вклада по 10% дали бы другую задачу.

Остаётся ещё вопрос: насколько полученное $\chi^2_{\min}$ типично для верной модели? Здесь требуется распределение минимума в ансамбле экспериментов. Оно отличается от распределения приращения $\Delta\chi^2$, используемого для интервалов. Разберём это в следующей главе.

Итоги главы

- Коррелированные измерения входят в МНК через полную ковариацию. Локальная обратная ковариация параметров равна половине матрицы вторых производных χ^2 в минимуме.
- Интервалы одного параметра строят по профилю, а совместные области требуют порога для соответствующего числа параметров. Вне линейной гауссовой задачи покрытие нужно обосновывать отдельно.
- Независимая гауссова калибровка даёт штрафной член. Линейное влияние мешающих параметров позволяет выполнить профилирование аналитически и получить эффективную ковариацию $V_y + HV_\eta H^T$.
- Общая нормировка меняет все предсказания согласованно. Подстановка отдельных измеренных значений вместо модельного масштаба может сместить результат; уже учтённую систематику не добавляют повторно после подгонки.

12.10. Задачи

Задача 1. Коррелированные измерения

Три гауссовых измерения Y_i имеют общее среднее μ и известную ковариацию

$$V = \sigma^2 \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}, \quad \sigma > 0, \quad -\frac{1}{2} < \rho < 1.$$

1. Запишите $\chi^2(\mu)$, найдите оценку $\hat{\mu}$ и её дисперсию.
2. Объясните, почему ρ меняет ошибку, но не меняет центральное значение.
3. Исследуйте пределы $\rho \rightarrow 0$, $\rho \rightarrow 1$ и $\rho \rightarrow -1/2$ изнутри допустимого интервала. Что происходит с собственными значениями V на границах?
4. Сравните истинную дисперсию оценки с ошибкой, которую вы сообщили бы при игнорировании корреляций.

Задача 2. Одна калибровка для многих измерений

Результаты описываются моделью $Y_i = \mu + \eta + \varepsilon_i$, $i = 1, \dots, N$, где ε_i независимы и имеют распределение $\mathcal{N}(0, s^2)$. Независимая калибровка даёт $Z \sim \mathcal{N}(\eta, \tau^2)$. Параметры μ, η фиксированы, дисперсии $s^2, \tau^2 > 0$ известны.

1. Запишите совместный $\chi^2(\mu, \eta)$ для полученных y_i и z .
2. Найдите общий минимум, профиль по μ и дисперсию оценки $\hat{\mu}$.
3. Проверьте эквивалентную запись с остатками $y_i - z - \mu$ и матрицей $V = s^2 I + \tau^2 \mathbf{1}\mathbf{1}^T$, где I — единичная матрица, $\mathbf{1}$ — вектор из единиц.
4. Что происходит с ошибкой $\hat{\mu}$ при $N \rightarrow \infty$? Что получилось бы, если ошибочно добавить τ^2 только к диагонали?
5. При $\mu = 10$, $\eta = 0.5$, $s = 1$, $\tau = 0.3$, $N = 20$ проверьте ответ псевдоэкспериментами. В каждом опыте заново генерируйте и основные данные, и калибровку.

Задача 3. Кривизна и нелинейная подгонка

Средний сигнал имеет вид $f(t) = Ae^{-\lambda t}$. Сгенерируйте пять независимых гауссовых измерений при $t_i = 0, 1, 2, 3, 4$ с, $A = 10$ мВ, $\lambda = 0.5$ с⁻¹ и известной одинаковой ошибке $s = 1$ мВ. Зафиксируйте начальное состояние генератора.

1. Подгоните $A > 0$ и $\lambda \geq 0$, используя несколько начальных точек.
2. В случае внутреннего минимума вычислите матрицу вторых производных χ^2 и локальную ковариацию. Проверьте множитель $1/2$.
3. Постройте контур $\Delta\chi^2 = 2.30$ и сравните с локальным эллипсом. Гарантирует ли само значение порога покрытие 68.3% в этой задаче?
4. Постройте профиль по λ и сечение с фиксированным $\hat{A}$.
5. Повторите расчёт при $s = 3$ мВ, сохранив те же стандартные гауссовы случайные числа. Исследуйте асимметрию профиля и влияние границы $\lambda = 0$.

Задача 4. Общая нормировка

Даны $y_1 = 8.00$ пб, $y_2 = 8.50$ пб и известные независимые гауссовы ошибки $s_1 = 0.16$ пб, $s_2 = 0.17$ пб. Модель имеет вид $Y_i = (1 + \alpha)\mu + \varepsilon_i$. Независимая калибровка $Z \sim \mathcal{N}(\alpha, f_N^2)$ дала $z = 0$ при $f_N = 0.10$.

1. Запишите совместный χ^2 , найдите минимум и аналитический профиль по μ .
2. Вычислите локальную ковариацию $(\hat{\mu}, \hat{\alpha})$ и границы $\Delta\chi_p^2 = 1$. Сравните их с $\hat{\mu} \pm \sigma_\mu$.
3. Постройте матрицу $V_{ij}(\mu) = s_i^2 \delta_{ij} + f_N^2 \mu^2$ и численно проверьте эквивалентность профильной записи при нескольких μ .
4. Повторите подгонку с фиксированной матрицей $V_{ij} = s_i^2 \delta_{ij} + f_N^2 y_i y_j$. Найдите веса и объясните сдвиг оценки.
5. Почему добавление $\ln \det V(\mu)$ меняет исходную профильную задачу?

Задача 5. Энергия по калибровке

Калибровка описывается прямой $y = a + bE$. Получены $\hat{a} = 2$ канала, $\hat{b} = 100$ каналов/МэВ, $\text{Var}(\hat{a}) = 28/3$ канал², $\text{Var}(\hat{b}) = 2$ канал²/МэВ² и $\text{cov}(\hat{a}, \hat{b}) = -4$ канал²/МэВ. Новое независимое измерение дало $y_* = 202$ канала с ошибкой $s_* = 2$ канала.

1. Найдите $\hat{E} = (y_* - \hat{a})/\hat{b}$.
2. Выведите её дисперсию в линейном приближении и вычислите ошибку в МэВ.
3. Повторите расчёт, отбросив ковариацию. Насколько изменился ответ?
4. Ковариация калибровки уже включает общую систематическую неопределённость. Нужно ли добавлять её вклад к ошибке энергии ещё раз?
5. При каких значениях $\hat{b}$ и его ошибки линейное приближение для отношения требует проверки?

Литература

- [1] D'Agostini, G. On the Use of the Covariance Matrix to Fit Correlated Data. *Nuclear Instruments and Methods in Physics Research Section A*, **346**, 306–311. 1994. doi:10.1016/0168-9002(94)90719-6.
- [2] Cowan, G. *Statistical Data Analysis*. Oxford University Press. 1998.