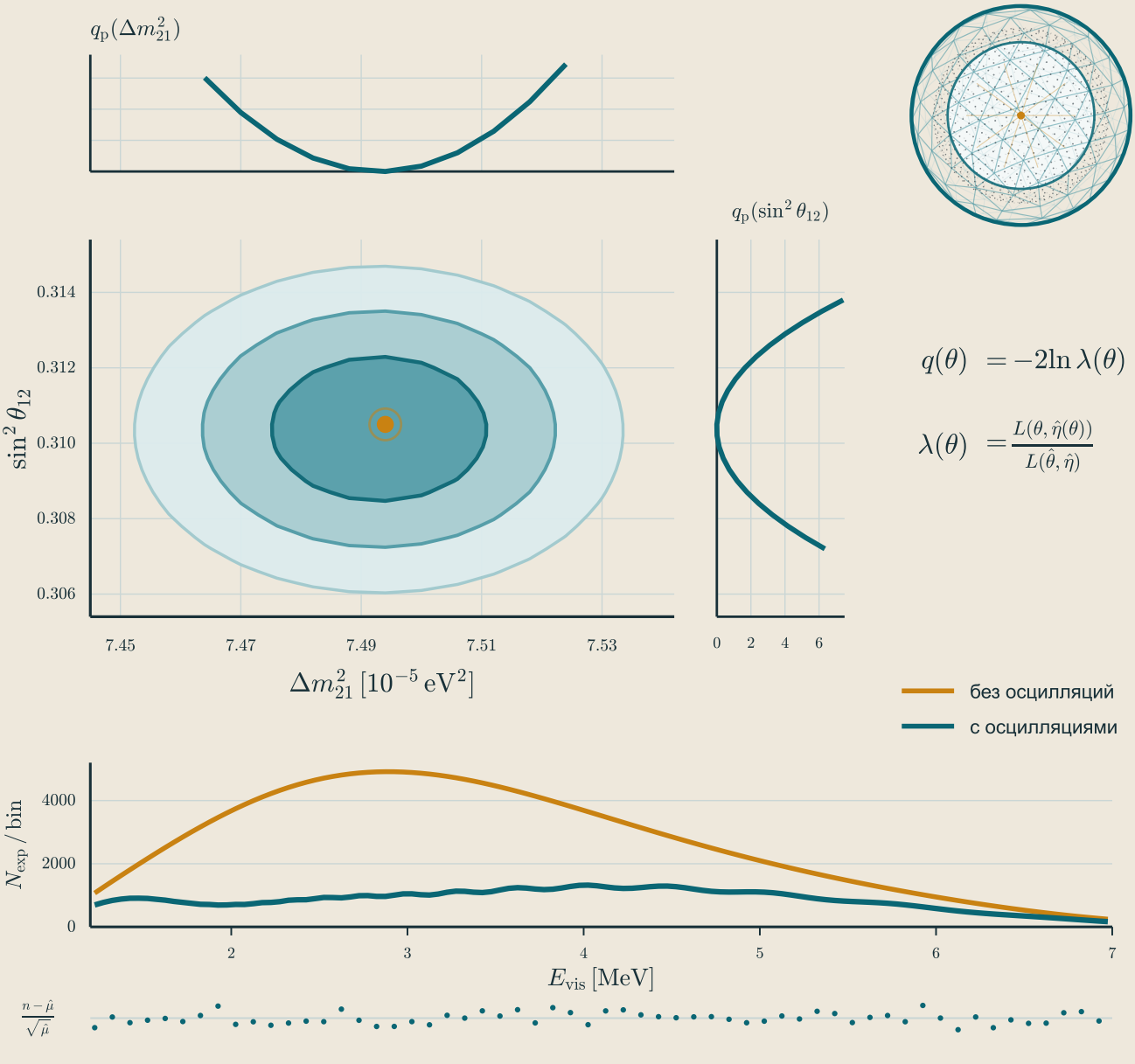
Статистические методы в физике

Практический курс для экспериментаторов и теоретиков

Дмитрий В. Наумов



Содержание

Книга посвящена статистическим методам, которые используются при анализе экспериментальных данных: от вероятностной модели и оценки параметров до проверки гипотез и формулировки физического результата.

Предисловие

Статистическая процедура входит в определение экспериментального результата. Выбор функции правдоподобия, учёт фона и систематических неопределённостей, построение доверительных интервалов и проверка гипотез определяют, какой физический вывод можно сделать по данным.

Книга выросла из курса по методам статистического анализа, которыми приходится пользоваться в экспериментальной физике. Материал отобран по опыту трёх десятилетий работы с данными в физике элементарных частиц, нейтринной физике и астрофизике.

Теоретическое изложение сопровождается физическими примерами, вычислениями, кодом на Python и интерактивными приложениями. Их можно использовать для проверки формул и для самостоятельного анализа простых моделей.

Для каждого метода обсуждаются статистическая модель, условия применимости и смысл полученного результата. Отдельное внимание уделено ошибкам, которые возникают при выборе модели, оценке параметров и интерпретации статистического вывода.

Курс создан в рамках программы «Физика нейтрино и астрофизика частиц». Он рассчитан на студентов старших курсов, аспирантов и исследователей, знакомых с основами теории вероятностей, математического анализа и программирования на Python.

РАЗДЕЛ КНИГИ

I. Вероятностный язык

Главы

Глава 1

Зачем физику статистика

Содержание

Аннотация

В этой главе объясняется, зачем физику нужен статистический анализ. На примерах из реальных научных статей мы познакомимся с первыми статистическими понятиями и проследим путь от теоретического расчёта и экспериментальных данных к физическому выводу.
Введём основные обозначения, обсудим частотную и байесовскую интерпретации вероятности — два принципиально разных взгляда на статистический вывод. Поговорим также о том, почему некоторые решения об анализе нужно принимать ещё до того, как исследователь увидел результат. Наконец, выполним первую подгонку по псевдоданным.

1.1. О чём этот курс?

В этом курсе мы разберём основные понятия, методы и язык, на котором физики обсуждают результаты эксперимента. Прочитайте несколько обычных фраз из научных статей:

- «Обнаружен сигнал со значимостью 5σ.»
- «Наблюдается значимый избыток событий.»
- «Область параметров исключена на уровне 95% C.L.»
- «Установлен верхний предел на сечение.»
- «Фит сошёлся.»
- «Доминирует систематическая неопределённость.»
- «Корреляции между бинами существенны.»
- «Мешающие параметры профилированы.»
- «Использована полная ковариационная матрица.»

Всё ли здесь понятно? Если нет, этот курс может оказаться полезным.

1.2. Четыре шага в научном исследовании

Зачем вообще физику статистика? Рассмотрим типичный путь экспериментального исследования.

1. **Теория и физическая модель.** Теория описывает физический процесс и позволяет вычислить наблюдаемые величины. Например, число и распределение по кинематическим переменным бозонов Хиггса, которые должны рождаться на БАК при заданных параметрах модели.
2. **Эксперимент.** Детектор регистрирует события и после реконструкции и отбора даёт конкретный набор данных $\mathbf{d}_{\text{obs}}$.
3. **Статистический анализ.** Данные $\mathbf{d}_{\text{obs}}$ сопоставляются с моделью. По ним оцениваются параметры, проверяются гипотезы и определяется статистическая неопределённость сделанного вывода.
4. **Физический результат.** Результат анализа формулируется как измерение параметров, наблюдение или открытие эффекта, исключение области параметров либо верхний предел. После этого он становится результатом научной статьи.

Статистический анализ связывает теоретический расчёт с физическим выводом из данных. Без него невозможно количественно ответить на главный вопрос эксперимента: насколько наблюдаемый результат согласуется с той

или иной физической моделью?
Заявления ветхозаветных старцев о том, что они всё видят на глазок и статистический анализ им не нужен, в современной экспериментальной физике имеют весьма ограниченную область применимости.

1.3. Обозначения

В этой книге и онлайн-курсе будем использовать следующие обозначения.

- Скалярные величины обозначаются курсивом: x, y, θ.
- Векторы — полужирным шрифтом:

$$\mathbf{x} = (x_1, x_2, \dots, x_n).$$

- Матрицы — полужирными заглавными буквами: $\mathbf{V}$, $\mathbf{M}$.
- Случайные величины обозначаются заглавными буквами: X, Y; случайные векторы — $\mathbf{X}$, $\mathbf{D}$.
- Их конкретные наблюдаемые значения обозначаются строчными буквами: x, y, $\mathbf{d}_{\text{obs}}$.
- Вероятность события A обозначается $P(A)$.
- Вероятность для дискретной величины или плотность вероятности для непрерывной будем обозначать $p(x)$ или $f(x)$.
- Параметр модели обозначается θ, а набор параметров — θ.
- Оценка параметра, вычисленная по данным, обозначается $\hat{\theta}$ или $\hat{\theta}$.
- Математическое ожидание: $\mathbb{E}[X]$.
- Дисперсия: $\mathrm{Var}(X)$.
- Стандартное отклонение:

$$\sigma_X = \sqrt{\mathrm{Var}(X)}.$$

- Ковариация: $\mathrm{cov}(X, Y)$.
- Коэффициент корреляции: $\rho(X, Y)$.

1.4. Что предсказывает теория?

Слово «предсказывает» звучит несколько странно — будто теория представляет собой дельфийского оракула. Это калька с английского *predict*, давно вошедшая в язык физиков. Имеется в виду просто результат вычисления.

Теория может предсказывать сечение процесса, спектр энергии частиц, вероятность распада или другую физическую величину. Но эксперимент наблюдает не непосредственно эти величины, а отклик реальной установки. Поэтому для статистического анализа нужна более полная модель, включающая физическое предсказание, фон, отклик детектора, эффективность регистрации и другие существенные детали эксперимента.

В результате мы приходим к вероятностной модели возможных данных:

$$p(\mathbf{d} \mid \theta),$$

где $\mathbf{d}$ обозначает возможный результат эксперимента, а θ — параметры модели.

Эта запись отвечает на вопрос:

> если параметры равны θ, какие данные $\mathbf{d}$ и с какой вероятностью может получить эксперимент?

Для дискретных данных $p(\mathbf{d} \mid \theta)$ означает вероятность, для непрерывных — плотность вероятности.

1.5. Что наблюдает эксперимент?

До проведения эксперимента будущие данные неизвестны. Их будем описывать случайным вектором $\mathbf{D}$ с распределением

$$p(\mathbf{d} \mid \theta).$$

Коротко это будем записывать как

$$\mathbf{D} \sim p(\mathbf{d} \mid \theta).$$

Знак $\sim$ здесь означает «распределён согласно».

После проведения эксперимента получен один конкретный результат:

$$\mathbf{d}_{\text{obs}}.$$

Таким образом, $\mathbf{D}$ — случайный результат ещё не проведённого или мысленно повторяемого эксперимента, а $\mathbf{d}_{\text{obs}}$ — уже наблюдаемые данные.

Под символом $\mathbf{d}$ может скрываться что угодно: число событий, набор измеренных энергий, гистограмма, времена регистрации или весь набор данных эксперимента.

1.6. Суть статистического анализа в двух словах

Итак, у нас есть наблюдаемые данные

$$\mathbf{d}_{\text{obs}}$$

и статистическая модель

$$p(\mathbf{d} \mid \theta).$$

Дальше возникают два основных типа вопросов.

1. **Оценивание параметров.** Какие значения параметров θ лучше всего согласуются с полученными данными? Результатом является оценка

   $$\hat{\theta}$$

   и её неопределённость.
2. **Проверка гипотез.** Насколько данные согласуются с определённой гипотезой? Например, совместимы ли они с отсутствием сигнала, можно ли исключить некоторую область параметров или достаточно ли велико наблюдаемое отклонение, чтобы говорить об открытии?

Обе задачи начинаются с одной и той же вероятностной модели

$$p(\mathbf{d} \mid \theta),$$

но отвечают на разные вопросы.

В этом курсе мы научимся формулировать эти вопросы точно и строить статистические процедуры, которые позволяют на них отвечать.

1.7. Два взгляда на вероятность

Если вам случится оказаться на большой конференции по статистике, вы быстро обнаружите представителей двух школ: частотной и байесовской.

Они используют одну и ту же теорему вероятностей, но по-разному понимают, к каким объектам вообще допустимо применять слово «вероятность».

1.7.1. Частотная вероятность

В частотном подходе вероятность связывают с относительной частотой события A в серии одинаковых опытов:

$$P(A) = \lim_{n\to\infty} \frac{\text{число наблюдений события } A}{n}. \tag{1.1}$$

Это операциональная интерпретация вероятности. Параметры модели считаются фиксированными, хотя их значения могут быть неизвестны. При повторении эксперимента меняются данные.

Например, если истинная масса частицы равна m, то при многократном повторении измерения будут меняться результаты эксперимента и построенные по ним оценки $\widehat{m}$, но сама масса m в этой постановке не флуктуирует.

1.7.2. Байесовская вероятность

Теорема Байеса для параметров модели записывается в виде

$$p(\theta \mid \mathbf{d}) = \frac{p(\mathbf{d} \mid \theta)\,\pi(\theta)}{p(\mathbf{d})}. \tag{1.2}$$

Здесь

- $p(\mathbf{d} \mid \theta)$ — вероятность или плотность данных при фиксированных параметрах модели;
- $\pi(\theta)$ — априорное распределение параметров;
- $p(\theta \mid \mathbf{d})$ — апостериорное распределение после получения данных;
- $p(\mathbf{d})$ — нормировочный множитель,

$$p(\mathbf{d}) = \int p(\mathbf{d} \mid \theta)\pi(\theta)\, d\theta.$$

БИОГРАФИЯ

Томас Байес

1701–1761

Томас Байес — английский пресвитер и математик. Его работу об обратной вероятности опубликовали после смерти автора. Формула, получившая имя Байеса, связывает условные вероятности и лежит в основе байесовского вывода.

Источник: MacTutor; фото MacTutor

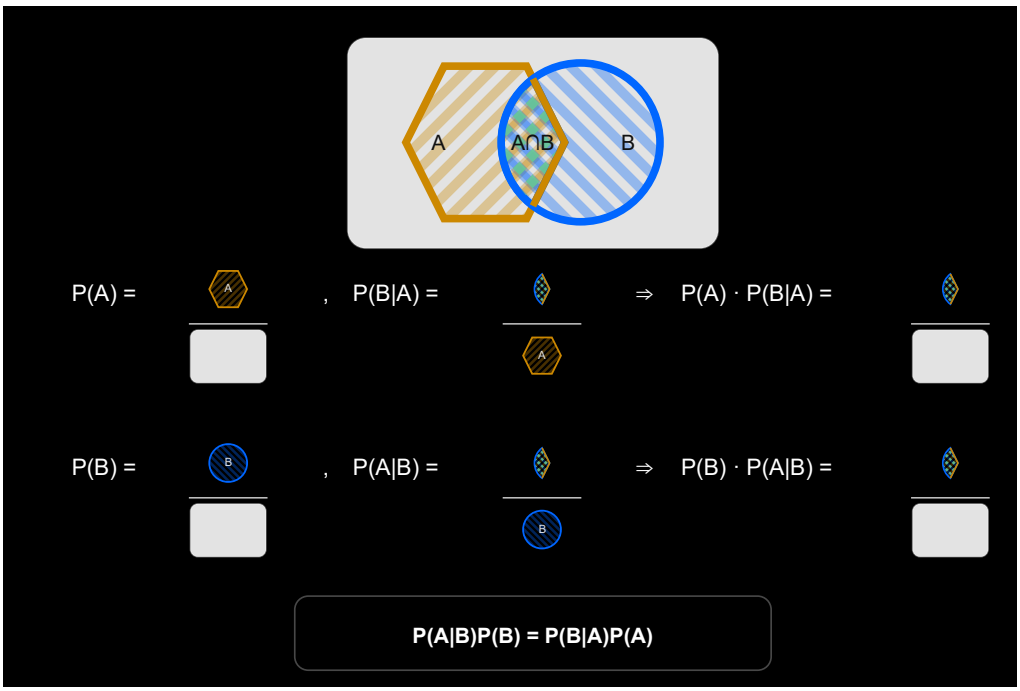


Рис. 1.1. *Визуальное доказательство теоремы Байеса*

На рисунке 1.1 обе части равенства

$$P(A \mid B)P(B) = P(B \mid A)P(A)$$

соответствуют одной и той же области пересечения $A \cap B$.

Сама формула возражений не вызывает. Вопрос начинается с множителя $\pi(\theta)$. Откуда известно распределение параметра до измерения?

У $p(\mathbf{d} \mid \theta)$ вполне прозрачный смысл. Фиксируем θ, многократно повторяем эксперимент и смотрим, как распределены данные $\mathbf{d}$. Именно такой опыт определяет вероятность в частотной постановке.

Попробуем тем же способом понять $\pi(\theta)$. Возьмём массу электрона. Чтобы буквально прочитать $\pi(m_e)$ как частоту, пришлось бы рассмотреть бесконечное число Вселенных с разными массами электрона и узнать, как эти массы распределены. В нашем распоряжении одна Вселенная и одна масса электрона. Никакой серии опытов, в которой природа каждый раз заново выбирает m_e, у нас нет.

В байесовской интерпретации $\pi(\theta)$ описывает степень знания о параметре, а не частоту его появления в разных Вселенных. Здесь и начинается спор. Слово «вероятность» применяется к двум разным объектам. В $p(\mathbf{d} \mid \theta)$ случайность относится к результату повторяемого эксперимента. В $\pi(\theta)$ она относится к нашему незнанию фиксированного параметра.

Последовательное байесовское обновление частично снимает эту проблему. Пусть сначала получены данные $\mathbf{d}_1$. Построенное по ним распределение $p(\theta \mid \mathbf{d}_1)$ можно использовать как априорное при анализе следующего независимого измерения $\mathbf{d}_2$:

$$p(\theta \mid \mathbf{d}_1, \mathbf{d}_2) \propto p(\mathbf{d}_2 \mid \theta)p(\theta \mid \mathbf{d}_1). \tag{1.3}$$

Апостериорное распределение первого эксперимента становится априорным для второго. Такое последовательное обновление однозначно определено, но вопрос о самом первом $\pi_0(\theta)$ остаётся.

Рассмотрим простейшее измерение $x \mid \theta \sim \mathcal{N}(\theta, s^2)$. Если на существенном для функции правдоподобия интервале взять плоское априорное распределение, то

после одного измерения

$$p(\theta \mid x) \propto \exp\left[-\frac{(x-\theta)^2}{2s^2}\right]. \tag{1.4}$$

После нормировки апостериорное распределение имеет ту же форму, что и функция правдоподобия. Плоское распределение можно получить как предел гауссова априорного распределения с очень большой шириной.

Теперь зададим другое начальное распределение $\pi_0(\theta)$. Для n независимых гауссовских измерений с одинаковой ошибкой s получим

$$p(\theta \mid x_1, \ldots, x_n) \propto \pi_0(\theta) \exp\left[-\frac{n(\theta-\bar{x})^2}{2s^2}\right]. \tag{1.5}$$

В уравнении 1.5 след первого выбора виден явно: $\pi_0(\theta)$ остаётся множителем при любом числе измерений. Когда функция правдоподобия становится узкой, а априорное распределение почти не меняется в этой узкой области, его влияние ослабевает. Если данных мало, априорное распределение быстро меняется в интересующей области или вообще запрещает часть значений параметра, след начального выбора сохраняется.

Тем не менее, байесовский подход имеет своих сторонников и он используется в большом числе физических анализов. Одним из важных практических преимуществ этого метода является более высокая скорость получения результата в задачах с большим числом параметров. После построения выборки из совместного апостериорного распределения маргинальные распределения отдельных параметров получаются простым суммированием по этой выборке. Для построения частотного профиля приходится при каждом значении интересующего параметра заново искать максимум по мешающим параметрам. Если параметров много и требуется несколько проекций, разница во времени может быть существенной.

Универсального выигрыша здесь нет. Получение многомерной апостериорной выборки само может оказаться сложнее серии минимизаций. Вычислительное удобство зависит от модели и алгоритма и ничего не говорит о физическом смысле $\pi(\theta)$.

Мне ближе частотная постановка. Основная часть книги и курса строится именно в таком подходе. Байесовский подход мы тоже разберём: читатель должен видеть исходное предположение и понимать, где оно вошло в результат.

1.7.3. Два разных интервала

Различие между частотным и байесовским подходами особенно хорошо видно на примере того, что в разговорной речи часто называют «интервалом ошибок».

Предположим, после измерения мы получили некоторый интервал для неизвестного параметра θ. Что означает число 95%?

Частотный подход отвечает на вопрос:

> Если много раз повторить эксперимент и каждый раз заново построить интервал, какая доля этих интервалов накроет истинное значение θ?

95%-й доверительный интервал — это процедура $C(\mathbf{D})$, для которой

$$P_\theta(\theta \in C(\mathbf{D})) = 0.95. \tag{1.6}$$

Случайным здесь является интервал: от эксперимента к эксперименту меняются данные $\mathbf{D}$ и построенные по ним границы. Параметр θ считается фиксированным.

После того как эксперимент уже выполнен и конкретный интервал получен, частотный подход не приписывает вероятности утверждению «θ находится внутри этого интервала».

Байесовский подход отвечает на другой вопрос:

> После получения этих данных как распределена наша неопределённость относительно самого параметра θ?

95%-й байесовский интервал $C_B(\mathbf{d})$ содержит 95% апостериорной вероятности:

$$\int_{C_B(\mathbf{d})} p(\theta \mid \mathbf{d})\, d\theta = 0.95. \tag{1.7}$$

Здесь уже можно сказать: при заданной модели и выбранном априорном распределении апостериорная вероятность того, что θ лежит в этом интервале, равна 95%.

Таким образом, одинаковое число 95% отвечает на два разных вопроса:

- в частотном подходе — **как ведёт себя процедура при повторении эксперимента?**
- в байесовском — **что мы знаем о параметре после получения данных?**

Границы этих интервалов иногда оказываются близкими или даже совпадают. Их статистический смысл от этого не становится одинаковым.

1.8. Как выглядят открытия и исключения

Теперь посмотрим на несколько реальных результатов из экспериментальной физики.

На первый взгляд это просто графики. Но именно из таких графиков физики делают выводы об открытии новых частиц, измерении параметров и исключении целых областей теории.

Попробуйте понять, что именно на них изображено и почему из этих данных следуют такие сильные утверждения.

1.8.1. Открытие бозона Хиггса

На рисунках 1.2 и 1.3 показаны результаты поисков бозона Хиггса в экспериментах ATLAS и CMS.

Перед чтением подписей попробуйте ответить:

- Что отложено по вертикальной оси?
- Что такое p-значение?
- Почему минимум кривой указывает на возможный сигнал?
- Что означают линии 1σ, 2σ, ..., 5σ?
- Почему достижение 5σ позволило говорить об открытии?
- Почему различают локальную и глобальную значимость?

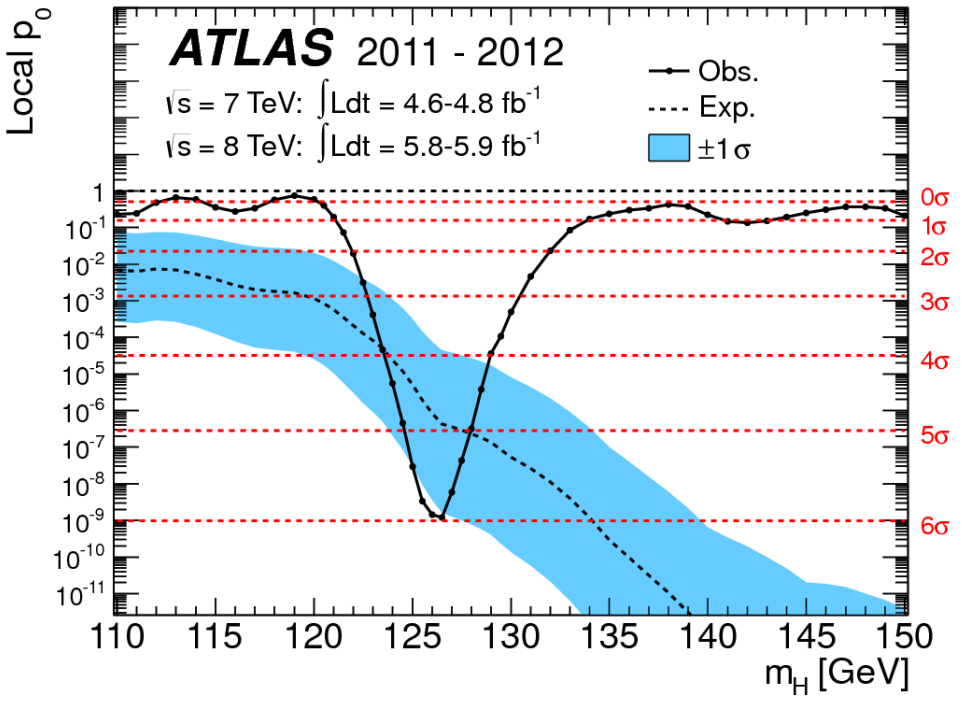


Рис. 1.2. *Локальное p-значение фоновой гипотезы в результате ATLAS*

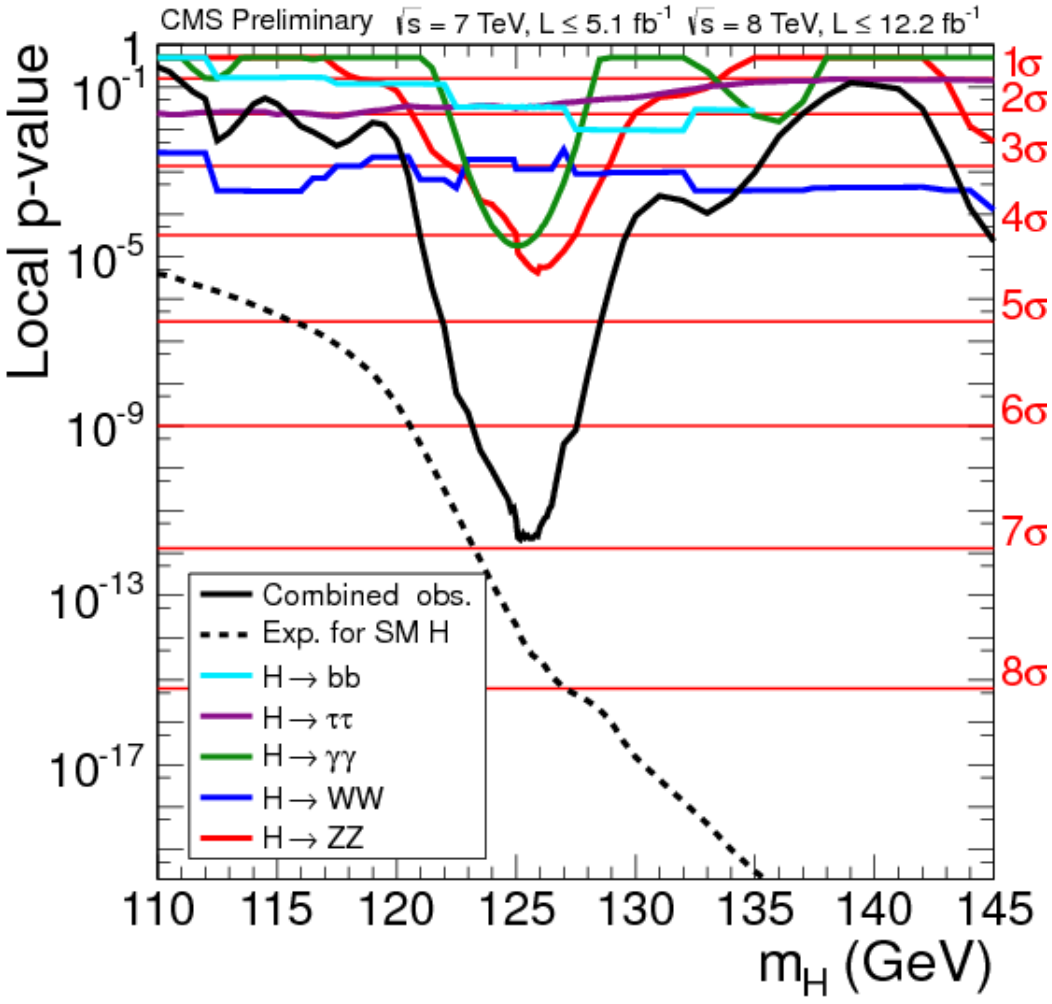


Рис. 1.3. *Локальное p-значение фоновой гипотезы в результате CMS*

По горизонтали перебирается предполагаемая масса бозона Хиггса. Чёрные кривые получены из экспериментальных данных. В районе 125 ГэВ оба эксперимента наблюдают особенно сильное отклонение от фоновой гипотезы.

ATLAS сообщил локальную значимость 5.9σ, CMS — 5.0σ [1, 2].

Почему именно эти числа означают открытие, мы разберём дальше.

Нобелевскую премию по физике 2013 года получили Франсуа Энглер и Питер Хиггс за теоретическое открытие механизма происхождения массы элементарных частиц. В формулировке Нобелевского комитета его экспериментальное подтверждение связано с открытием новой частицы экспериментами ATLAS и CMS [3].

1.8.2. Измерение параметра в Daya Bay

Другой тип результата показан на рисунке 1.4. Эксперимент Daya Bay измерял угол смешивания нейтрино θ_{13}.

Попробуйте ответить:

- Что означает красная кривая?
- Что такое *best fit*?
- Почему минимум χ^2 определяет измеренное значение параметра?
- Что именно означает значимость 5.2σ?
- Означает ли хороший *best fit*, что теория доказана?

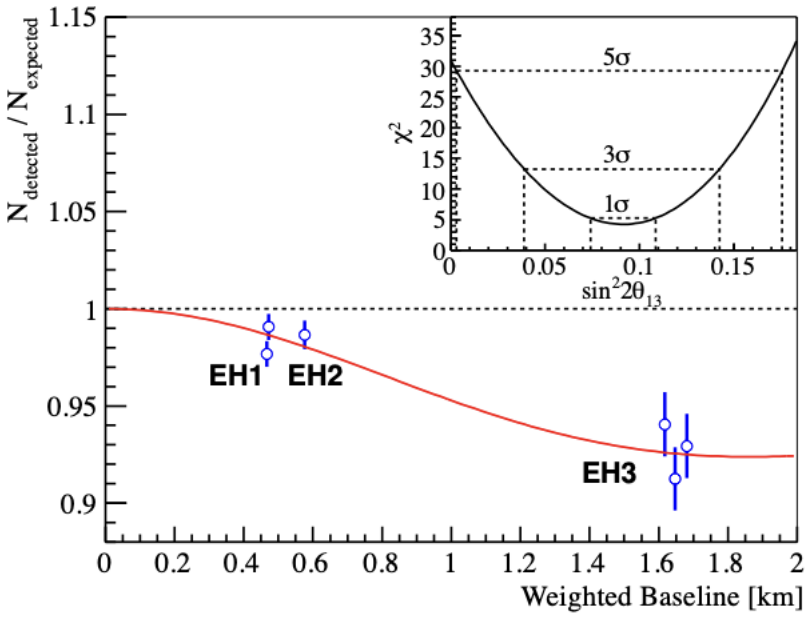


Рис. 1.4. *Измерение ненулевого угла смешивания θ_{13} в Daya Bay*

Красная кривая показывает осцилляционное предсказание при значении $\sin^2 2\theta_{13}$, найденном из совместной подгонки данных. Во вставке показана зависимость χ^2 от этого параметра.

Daya Bay отверг гипотезу $\theta_{13} = 0$ со значимостью 5.2σ [4].

Почему минимум χ^2 даёт оценку параметра и откуда возникает статистическая значимость, пока оставим без ответа.

1.8.3. Исключение области параметров

Наконец, посмотрим на результат поиска стерильных нейтрино в Daya Bay.

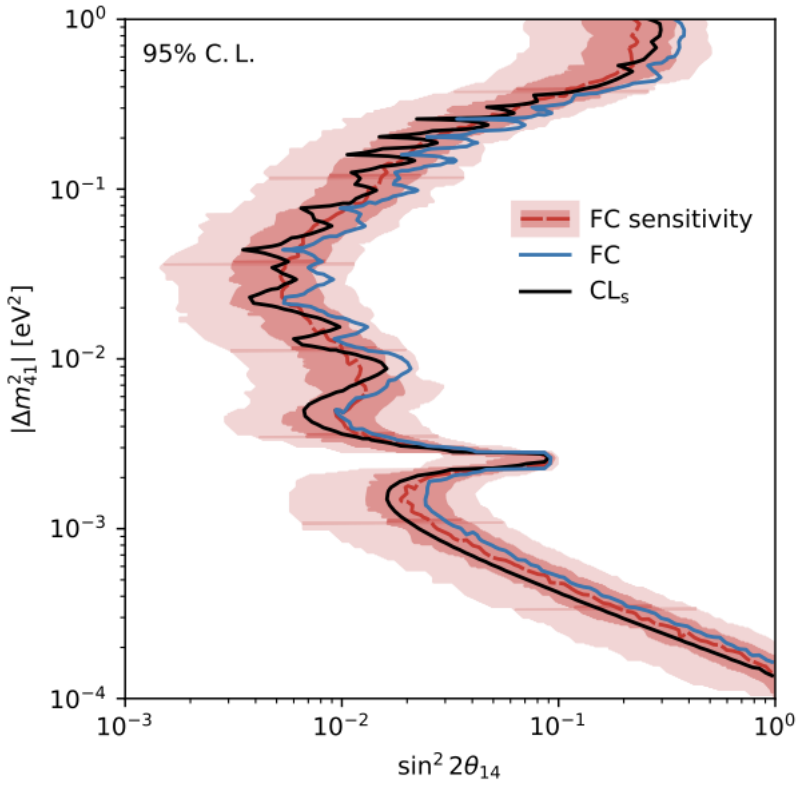


Рис. 1.5. *Исключение области параметров стерильных нейтрино в Daya Bay*

Здесь уже нет одного измеренного числа. Результат представлен областью на плоскости параметров

$$\sin^2 2\theta_{14}, \qquad |\Delta m^2_{41}|.$$

Попробуйте понять:

- Что означает фраза «область исключена на уровне 95%»?
- Почему результатом является контур, а не одна точка?
- Что означает *sensitivity*?
- Почему на одном графике показаны две разные статистические процедуры — Feldman–Cousins и CL_s?
- Почему их границы не обязаны совпадать?

Полный набор Daya Bay содержит $5.55 \cdot 10^6$ кандидатов за 3158 дней [5].

К концу курса все элементы этих четырёх рисунков должны стать понятными.

1.9. Первый фит: ширина гауссианы

Рассмотрим выборку из гауссова распределения с известным центром $x_0 = 0$ и неизвестной шириной σ:

$$f(x;\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\!\left(-\frac{x^2}{2\sigma^2}\right). \tag{1.8}$$

Сгенерируем N псевдособытий при некотором фиксированном значении σ_{true} и разобьём ось x на интервалы. Для интервала i модель предсказывает среднее число событий

$$\mu_i(\sigma) = N \int_{\text{бин } i} f(x;\sigma)\,dx,$$

а в сгенерированной выборке наблюдается n_i событий.

Теперь нужно определить, какое значение σ лучше всего описывает полученную гистограмму. Для этого используем пуассоновский девианс — статистику отношения правдоподобий:

$$\chi^2(\sigma) = 2 \sum_i \left[\mu_i(\sigma) - n_i + n_i \ln \frac{n_i}{\mu_i(\sigma)} \right]. \tag{1.9}$$

Для пустого интервала, $n_i = 0$, последнее слагаемое по пределу равно нулю.

Чем меньше $\chi^2(\sigma)$, тем лучше модель при данном σ согласуется с выборкой. Значение в точке минимума определяет оценку

$$\hat{\sigma} = \arg\min_{\sigma} \chi^2(\sigma).$$

1.9.1. Интерактив: найдите ширину распределения

Сначала попробуйте подобрать σ на глаз по гистограмме. Затем включите результат фита и сравните свою оценку с $\hat{\sigma}$ и истинным значением σ_{true}.

Интерактив. Подбор ширины распределения
Отсканируйте QR-код, чтобы открыть интерактив и самостоятельно подобрать ширину распределения.

При новой генерации псевдоданных меняется положение минимума и, следовательно, оценка $\hat{\sigma}$. Истинный параметр σ_{true} при этом остаётся тем же. С ростом числа событий кривая $\chi^2(\sigma)$ обычно становится уже: данные точнее определяют ширину распределения.

Итоги главы

- Статистическая модель $p(\mathbf{d} \mid \theta)$ задаёт распределение возможных данных при фиксированных параметрах модели. Наблюдаемые данные $\mathbf{d}_{\text{obs}}$ — одна конкретная реализация этого распределения.
- В статистическом анализе возникают две разные задачи: оценивание параметров и проверка гипотез.

- Параметр модели и его оценка — разные объекты. При повторении эксперимента параметр остаётся фиксированным, а вычисленная по данным оценка $\hat{\theta}$ меняется.
- Частотный и байесовский подходы по-разному интерпретируют вероятность и статус неизвестных параметров. В байесовском выводе результат зависит также от выбранного априорного распределения.
- Измерение параметра, открытие сигнала и исключение области параметров — разные статистические выводы. Их смысл определяется не только данными, но также моделью и выбранной статистической процедурой.
- Уже в простейшей подгонке видно, как из случайных данных возникает оценка параметра: истинная ширина σ_{true} фиксирована, а найденное значение $\hat{\sigma}$ флуктуирует между выборками.

1.10. Задачи

Задача 1. Модель и данные

В некотором эксперименте теория предсказывает число событий

$$\mathbf{t} = (10,\ 20,\ 10).$$

Эксперимент дал результат

$$\mathbf{d} = (12,\ 18,\ 11).$$

Ответьте:

1. Что здесь является предсказанием модели?
2. Что является данными эксперимента?
3. В каких бинах данные больше предсказания?
4. В каких бинах данные меньше предсказания?
5. Можно ли по этим числам сразу сказать, что теория неверна?

Задача 2. Случайная величина и измеренное значение

До измерения результат эксперимента неизвестен. Обозначим его случайной величиной X.
После измерения получено значение

$$x = 7.3.$$

Ответьте:

1. Что обозначает X?
2. Что обозначает x?
3. Почему до измерения мы говорим о X, а после измерения — о x?
4. Приведите свой пример случайной величины и её измеренного значения.

Задача 3. Частота как оценка вероятности

Детектор зарегистрировал 100 событий. Из них 23 события прошли отбор.
Ответьте:

1. Какую долю событий отобрал алгоритм?
2. Как по этим данным оценить вероятность того, что следующее событие пройдёт отбор?
3. Будет ли эта оценка точно равна истинной вероятности?
4. Что изменится, если вместо 100 событий было бы зарегистрировано 10 000 событий?

Задача 4. Условная вероятность без формул

В выборке есть 1000 событий:

- 100 сигнальных событий;
- 900 фоновых событий.

Алгоритм помечает событие как «интересное». Он пометил:

- 80 сигнальных событий;
- 90 фоновых событий.

Ответьте:

1. Сколько всего событий помечено как «интересные»?
2. Какая доля сигнальных событий была помечена?
3. Какая доля помеченных событий действительно является сигналом?
4. Почему это разные числа?
5. Что важнее для утверждения «мы нашли сигнал»: доля помеченных сигнальных событий или доля сигнала среди помеченных событий?

Задача 5. Точка наилучшего согласия

Одна и та же форма кривой сравнивается с данными при разных значениях параметра σ.
Получены такие значения меры расхождения:

σ	расхождение
0.6	18.4
0.8	7.2
1.0	2.1
1.2	3.5
1.4	9.8

Ответьте:

1. При каком значении σ модель лучше всего согласуется с данными?
2. Почему это значение называется точкой наилучшего согласия?
3. Значит ли это, что истинное значение σ точно равно найденному?
4. Что нужно знать дополнительно, чтобы говорить об ошибке найденного значения?

Задача 6. Флуктуация или систематический сдвиг?

Ожидаемое число событий в каждом запуске равно примерно 100.
В первом случае получены результаты:

$$98, \quad 103, \quad 99, \quad 101, \quad 100.$$

Во втором случае получены результаты:

$$113, \quad 111, \quad 115, \quad 112, \quad 114.$$

Ответьте:

1. В каком случае данные выглядят как случайные флуктуации около 100?
2. В каком случае виден возможный систематический сдвиг?
3. Почему отдельное значение 103 не выглядит подозрительным?
4. Почему последовательность значений около 113 уже выглядит подозрительной?
5. Что следовало бы проверить в эксперименте во втором случае?

Задача 7. Следы априорного распределения

Пусть измерения x_i независимы и имеют гауссовское распределение

$$x_i \mid \theta \sim \mathcal{N}(\theta, 1).$$

Рассмотрите выборку со средним $\bar{x} = 1$ сначала при $n = 1$, а затем при $n = 100$. Сравните три априорных распределения:

$$\pi_1(\theta) \propto 1, \qquad \pi_2(\theta) \propto \exp\left(-\frac{\theta^2}{2}\right), \qquad \pi_3(\theta) \propto \frac{1}{1+\theta^2}.$$

Для каждого случая постройте функцию

$$p(\theta \mid x_1, \dots, x_n) \propto \pi_k(\theta) \exp\left[-\frac{n(\theta - \bar{x})^2}{2}\right].$$

Ответьте:

1. Насколько различаются три результата после одного измерения?
2. Что меняется после ста измерений?
3. При каком условии влияние априорного распределения становится малым?
4. Что произойдёт, если оно равно нулю в окрестности $\bar{x}$?
5. Устраняет ли накопление данных вопрос о выборе самого первого априорного распределения?

Литература

[1] 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.
[2] 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.
[3] The Royal Swedish Academy of Sciences. The Nobel Prize in Physics 2013: Press Release. 2013. https://www.nobelprize.org/prizes/physics/2013/press-release/.
[4] Daya Bay Collaboration. Observation of Electron-Antineutrino Disappearance at Daya Bay. *Physical Review Letters*, **108**, 171803. 2012. doi:10.1103/PhysRevLett.108.171803.
[5] Daya Bay Collaboration. Search for a Light Sterile Neutrino with the Full Configuration of the Daya Bay Experiment. *Physical Review Letters*, **133**, 051801. 2024. doi:10.1103/PhysRevLett.133.051801. https://arxiv.org/abs/2404.01687.

Глава 2

Случайные величины, распределения, моменты

Содержание

Аннотация

В этой главе мы подробнее разберём случайные величины и их распределения. Введём функцию вероятностей для дискретной величины, плотность вероятности и кумулятивную функцию распределения для непрерывной. Затем определим моменты и на примерах выведем среднее и дисперсию для биномиального, пуассоновского и нормального распределений.

2.1. Случайность и случайные величины

Откуда в физических данных возникает случайность? В классической физике её обычно связывают с неполным знанием начальных условий и большим числом степеней свободы. Брошенная монета движется по законам механики, но для предсказания результата нужно было бы знать её начальное положение, скорость, вращение, сопротивление воздуха и ещё множество деталей.

В квантовой физике вероятностный характер заложен в саму теорию. Даже если начальное состояние приготовлено полностью, теория предсказывает вероятности возможных результатов измерения.

Для статистического анализа в обоих случаях эксперимент даёт одну конкретную реализацию случайного процесса. При повторении опыта результат меняется, и это изменение должно быть описано вероятностной моделью.

В физике **случайной величиной** X будем называть экспериментально измеряемую величину, значение которой нельзя знать до проведения эксперимента. В математической формулировке X каждому возможному исходу эксперимента сопоставляет число [1].

Например, N может обозначать число нейтрино, зарегистрированных детектором за сутки. До начала измерения N — случайная величина. После его окончания мы получим определённое число, скажем

$$n = 17.$$

Чтобы описать N, нужно указать все возможные значения и вероятности их получения. Вместе они образуют **распределение случайной величины**.

Если X принимает отдельные значения x_k, распределение задаётся **функцией вероятностей** (*probability mass function*, PMF):

$$p_k = P(X = x_k), \qquad \sum_k p_k = 1. \tag{2.1}$$

Для непрерывной случайной величины вводится **плотность вероятности** (*probability density function*, PDF):

$$P(a \le X < b) = \int_a^b f(x)\,dx, \qquad \int_{-\infty}^{+\infty} f(x)\,dx = 1. \tag{2.2}$$

Сумма в формуле 2.1 и второй интеграл в формуле 2.2 выражают условие нормировки: один из возможных результатов обязательно произойдёт.

Вероятность для непрерывной величины определяется интегралом плотности. Поэтому $P(X = x) = 0$, а вероятность попадания в конечный интервал может быть ненулевой. Для короткого интервала ширины Δx эту вероятность можно записать как $P(x \le X < x + \Delta x) \approx f(x)\,\Delta x$. Если x измеряется в МэВ, плотность $f(x)$ имеет размерность МэВ$^{-1}$, и произведение $f(x)\,\Delta x$ остаётся безразмерным.

2.1.1. Кумулятивная функция распределения (CDF)

Нас обычно интересует вероятность получить результат меньше некоторого значения a. Для этого вводится **кумулятивная функция распределения**, или CDF (*cumulative distribution function*):

$$F(a) = P(X < a). \tag{2.3}$$

Для непрерывной случайной величины она равна интегралу плотности:

$$F(a) = \int_{-\infty}^{a} f(x)\,dx. \tag{2.4}$$

При движении a слева направо функция $F(a)$ монотонно растёт от нуля до единицы. Если плотность непрерывна, то

$$\frac{dF(a)}{da} = f(a). \tag{2.5}$$

Вероятность попадания в интервал $[a, b)$ теперь вычисляется одной строкой:

$$P(a \le X < b) = F(b) - F(a). \tag{2.6}$$

Для непрерывного распределения со строго возрастающей F **квантиль** x_q определяется условием

$$F(x_q) = q. \tag{2.7}$$

Например, медиана $x_{0.5}$ делит такое распределение пополам: вероятность получить значение слева или справа от неё равна $1/2$. Квантили понадобятся нам при построении доверительных интервалов.

2.2. Средние, дисперсии и моменты

Распределение содержит всю вероятностную информацию о случайной величине. На практике часто хочется начать с нескольких чисел: где расположен центр распределения и насколько широко разбросаны его значения.

2.2.1. Математическое ожидание

Для произвольной функции $g(X)$ математическое ожидание определяется формулой

$$\mathbb{E}[g(X)] = \begin{cases} \displaystyle\sum_{k} g(x_k)\,p_k, & X \text{ — дискретная величина,} \\ \displaystyle\int_{-\infty}^{+\infty} g(x)\,f(x)\,dx, & X \text{ — непрерывная величина.} \end{cases} \tag{2.8}$$

Положив $g(X) = X^n$, получим n-й **начальный момент** распределения:

$$\langle x^n \rangle = \mathbb{E}[X^n]. \tag{2.9}$$

Первый момент — это среднее значение

$$\mu = \langle x \rangle = \mathbb{E}[X]. \tag{2.10}$$

Формула напоминает определение центра масс. Если представить $f(x)$ как плотность единичной массы на числовой оси, то точка μ действительно будет её центром масс.

2.2.2. Дисперсия

Среднее указывает центр распределения, но ничего не говорит о его ширине. Для описания разброса используется **дисперсия**:

$$\begin{aligned}\operatorname{Var}(X) &= \mathbb{E}\big[(X-\mu)^2\big] \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \langle x^2\rangle - \langle x\rangle^2.\end{aligned} \tag{2.11}$$

Квадрат не позволяет положительным и отрицательным отклонениям взаимно уничтожиться. Корень из дисперсии называется **стандартным отклонением**:

$$\sigma_X = \sqrt{\operatorname{Var}(X)}. \tag{2.12}$$

Стандартное отклонение имеет ту же размерность, что и сама величина X.

Для функции $g(X)$ определение остаётся тем же:

$$\operatorname{Var}(g(X)) = \mathbb{E}\Big[\big(g(X) - \mathbb{E}[g(X)]\big)^2\Big]. \tag{2.13}$$

Все эти формулы предполагают, что соответствующие сумма или интеграл сходятся. У распределения Коши, которое мы встретим в следующей главе, математического ожидания и дисперсии нет. Более того, даже полный набор моментов не всегда определяет распределение однозначно. Поэтому моменты полезны, но не заменяют саму функцию распределения.

2.3. Базовые распределения

В этой главе нам понадобятся три распределения.

- Биномиальное распределение описывает число успехов в фиксированном числе испытаний.
- Распределение Пуассона описывает число событий за фиксированное время или в фиксированной области пространства.
- Нормальное распределение описывает непрерывную величину и часто возникает при сложении большого числа малых независимых вкладов.

Разберём их по порядку.

2.4. Счёт успехов: биномиальная модель

Предположим, что в эксперименте проводится n независимых испытаний. В каждом из них возможны два исхода, которые условно назовём успехом и неудачей. Вероятность успеха равна p и остаётся одной и той же во всех испытаниях.

Обозначим через K полное число успехов. Вероятность любой определённой последовательности из k успехов и $n-k$ неудач равна

$$p^k (1-p)^{n-k}.$$

Успехи можно расположить среди n испытаний

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

различными способами. Поэтому

Биномиальное распределение задаётся функцией вероятностей

$$P(K = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n. \tag{2.14}$$

Короткая запись имеет вид

$$K \sim \mathrm{Bin}(n, p).$$

Условие нормировки следует непосредственно из бинома Ньютона:

$$\sum_{k=0}^{n} P(K = k) = \bigl[p + (1-p)\bigr]^n = 1.$$

2.4.1. Пример биномиального распределения

Пусть зарегистрировано n распадов W-бозона. Каждый распад либо происходит по каналу $W \to \mu\nu$, либо нет. Если вероятность этого канала равна p, то число распадов по данному каналу имеет биномиальное распределение.

На рисунке 2.1 показаны три распределения при $p = 0.3$ и $n = 5$, 20 и 100.

```
from math import comb
import numpy as np

p = 0.3
for n in [5, 20, 100]:
    k = np.arange(n + 1)
    pmf = np.array([comb(n, i) * p**i * (1-p)**(n-i) for i in k])
```

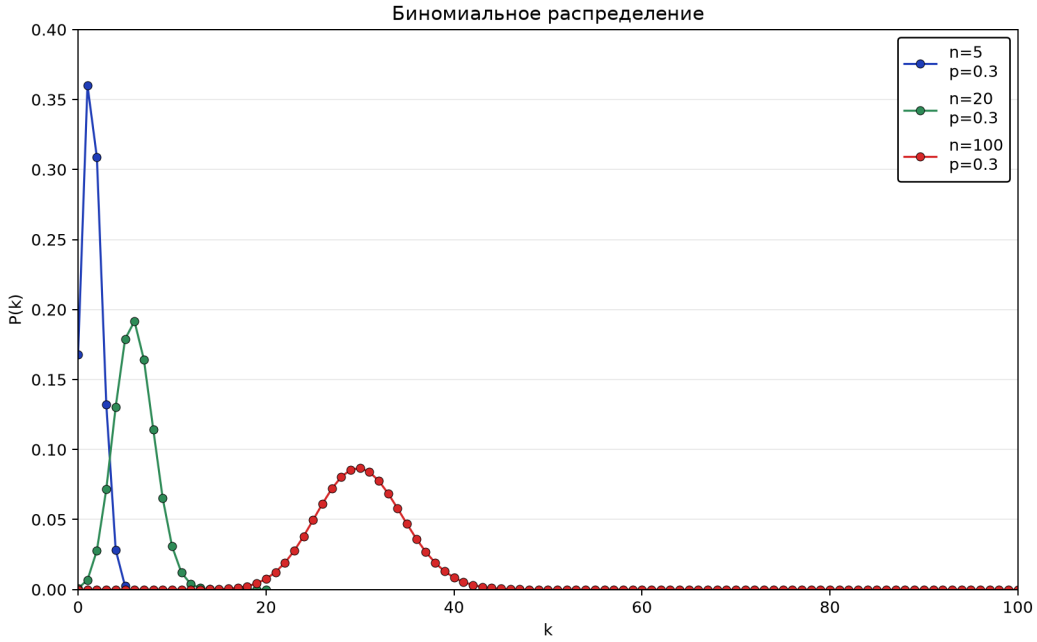


Рис. 2.1. *Биномиальные распределения при $p = 0.3$ и трёх значениях числа испытаний: $n = 5$, 20 и 100*

С ростом n максимум смещается вслед за средним значением np. Абсолютная ширина распределения растёт как $\sqrt{np(1-p)}$, а относительная ширина уменьшается

как $1/\sqrt{n}$. При больших np и $n(1-p)$ форма становится близкой к гауссовой, хотя величина K по-прежнему принимает только целые значения.

2.4.2. Моменты биномиального распределения

Для биномиального распределения

$$\mathbb{E}[K] = np, \qquad \operatorname{Var}(K) = np(1-p). \tag{2.15}$$

Выведем обе формулы. Этот вывод заодно показывает полезный приём с факториальными моментами.

Вывод

2.4.2.1. Среднее

Начнём с определения математического ожидания:

$$\mathbb{E}[K] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}.$$

Используем тождество

$$k\binom{n}{k} = n\binom{n-1}{k-1}.$$

Тогда

$$\begin{aligned} \mathbb{E}[K] &= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1}(1-p)^{n-k} \\ &= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^{j}(1-p)^{n-1-j} \\ &= np. \end{aligned} \tag{2.16}$$

Последняя сумма равна единице по биному Ньютона.

Вывод

2.4.2.2. Дисперсия

Для вычисления дисперсии удобно начать с факториального момента $K(K-1)$:

$$\mathbb{E}[K(K-1)] = \sum_{k=0}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k}.$$

Поскольку

$$k(k-1)\binom{n}{k} = n(n-1)\binom{n-2}{k-2},$$

тот же приём даёт

$$\mathbb{E}[K(K-1)] = n(n-1)p^2. \tag{2.17}$$

Обычный второй момент связан с факториальным тождеством

$$K^2 = K(K-1) + K.$$

Поэтому

$$\begin{aligned}\operatorname{Var}(K) &= \mathbb{E}[K^2] - \mathbb{E}[K]^2 \\ &= n(n-1)p^2 + np - n^2p^2 \\ &= np(1-p).\end{aligned} \tag{2.18}$$

2.5. Счёт редких событий: пуассоновская модель

Теперь рассмотрим другой эксперимент. Радиоактивный источник работает в течение заданного времени Δt, а детектор считает распады. Число испытаний заранее не фиксировано. Задана средняя частота регистрации r, и среднее число событий за время наблюдения равно

$$\lambda = r\,\Delta t. \tag{2.19}$$

Если интенсивность не меняется со временем, а события происходят независимо, число зарегистрированных событий N подчиняется распределению Пуассона.

Распределение Пуассона задаётся функцией вероятностей

$$P(N = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \qquad k = 0, 1, 2, \dots, \tag{2.20}$$

где $\lambda > 0$ — среднее число событий. Короткая запись имеет вид

$$N \sim \mathrm{Pois}(\lambda).$$

БИОГРАФИЯ

Симеон Дени Пуассон

1781–1840

Симеон Дени Пуассон — французский математик и физик. Он работал над теорией вероятностей, механикой, теорией упругости, электричеством и магнетизмом. Распределение, носящее его имя, появилось в исследовании вероятностей судебных решений и сегодня постоянно используется при подсчёте событий в физическом эксперименте.

Источник: MacTutor; фото MacTutor

2.5.1. Пример распределения Пуассона

На рисунке 2.2 показаны распределения Пуассона при $\lambda = 1$, 4 и 10.

```
from math import factorial
import numpy as np

for lam in [1, 4, 10]:
    k = np.arange(26)
    pmf = np.exp(-lam) * lam**k / np.array([factorial(i) for i in k])
```

При $\lambda = 1$ основная вероятность сосредоточена около нуля и распределение заметно асимметрично. С ростом λ максимум смещается вправо, ширина увеличивается, а форма становится всё более симметричной. При больших λ распределение Пуассона можно приближать нормальным распределением. Условия

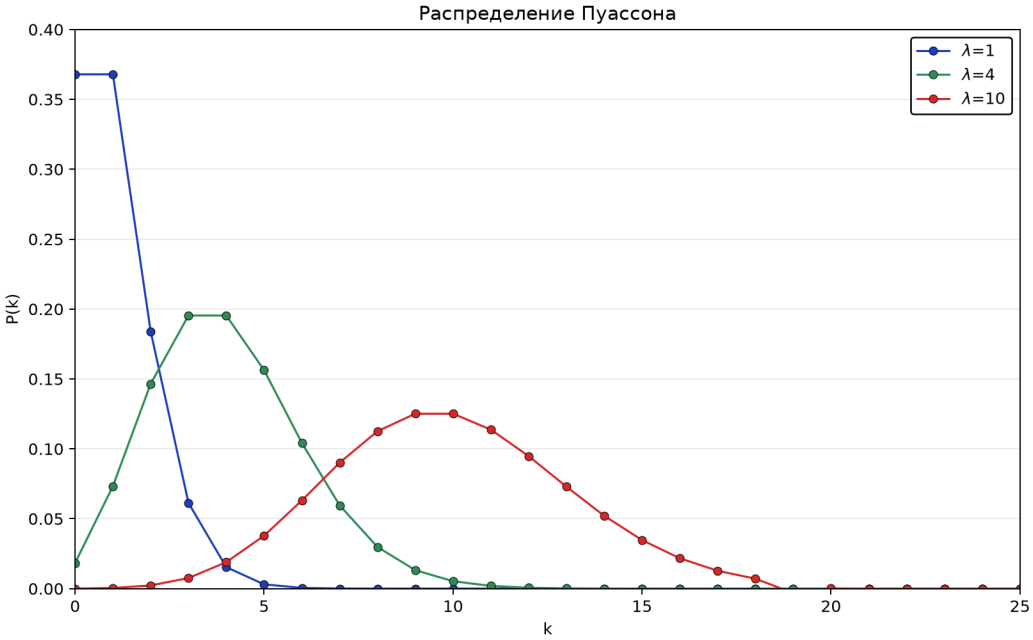


Рис. 2.2. *Распределения Пуассона при среднем числе событий* $\lambda = 1, 4$ *и* 10

и точность такого приближения мы обсудим в главе о центральной предельной теореме.

2.5.2. Моменты распределения Пуассона

У распределения Пуассона среднее и дисперсия равны одному и тому же параметру:

$$\mathbb{E}[N] = \lambda, \qquad \mathrm{Var}(N) = \lambda. \tag{2.21}$$

Вывод

2.5.2.1. Среднее

По определению математического ожидания

$$\mathbb{E}[N] = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!}.$$

Слагаемое при $k = 0$ равно нулю. В остальных слагаемых вынесем один множитель λ и сдвинем индекс $j = k - 1$:

$$\begin{aligned} \mathbb{E}[N] &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\ &= \lambda e^{-\lambda} e^{\lambda} = \lambda. \end{aligned} \tag{2.22}$$

Вывод

2.5.2.2. Дисперсия

Снова начнём со второго факториального момента:

$$\mathbb{E}[N(N-1)] = \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k e^{-\lambda}}{k!}.$$

Первые два слагаемых равны нулю, а при $k \geq 2$

$$\frac{k(k-1)}{k!} = \frac{1}{(k-2)!}.$$

Отсюда

$$\begin{aligned} \mathbb{E}[N(N-1)] &= \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} \\ &= \lambda^2 e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\ &= \lambda^2. \end{aligned} \tag{2.23}$$

Используя $N^2 = N(N-1) + N$, получаем

$$\begin{aligned} \operatorname{Var}(N) &= \mathbb{E}[N^2] - \mathbb{E}[N]^2 \\ &= \lambda^2 + \lambda - \lambda^2 \\ &= \lambda. \end{aligned} \tag{2.24}$$

2.6. Непрерывные ошибки: нормальная модель

Биномиальное и пуассоновское распределения описывают целое число событий. Результат измерения энергии, времени или координаты обычно рассматривается как непрерывная величина. Самая известная модель для такой величины — нормальное, или гауссово, распределение.

Нормальное распределение с параметрами μ и $\sigma > 0$ имеет плотность

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \tag{2.25}$$

Здесь μ — среднее значение, а σ — стандартное отклонение. Короткая запись имеет вид

$$X \sim \mathcal{N}(\mu, \sigma^2).$$

Почему это распределение называется нормальным? Карл Пирсон, который ввёл это название в широкое употребление, позднее заметил его неудачную сторону:

> Много лет назад я назвал кривую Лапласа-Гаусса нормальной кривой. Это название, избегая международного спора о приоритете, имеет недостаток: оно заставляет думать, что все прочие распределения частот в том или ином смысле «ненормальны».
>
> К. Пирсон, 1920 [2]

Причина частого появления гауссовой формы связана с центральной предельной теоремой. Сумма большого числа малых независимых вкладов при достаточно общих условиях приближается нормальным распределением. Этому будет посвящена отдельная глава.

2.6.1. Пример нормального распределения

На рисунке 2.3 показаны три плотности с параметрами $(\mu, \sigma) = (0, 1), (0, 2)$ и $(2, 1)$.

```
import numpy as np

x = np.linspace(-6, 6, 400)
for mu, sigma in [(0, 1), (0, 2), (2, 1)]:
    pdf = np.exp(-0.5*((x-mu)/sigma)**2)/(sigma*np.sqrt(2*np.pi))
```

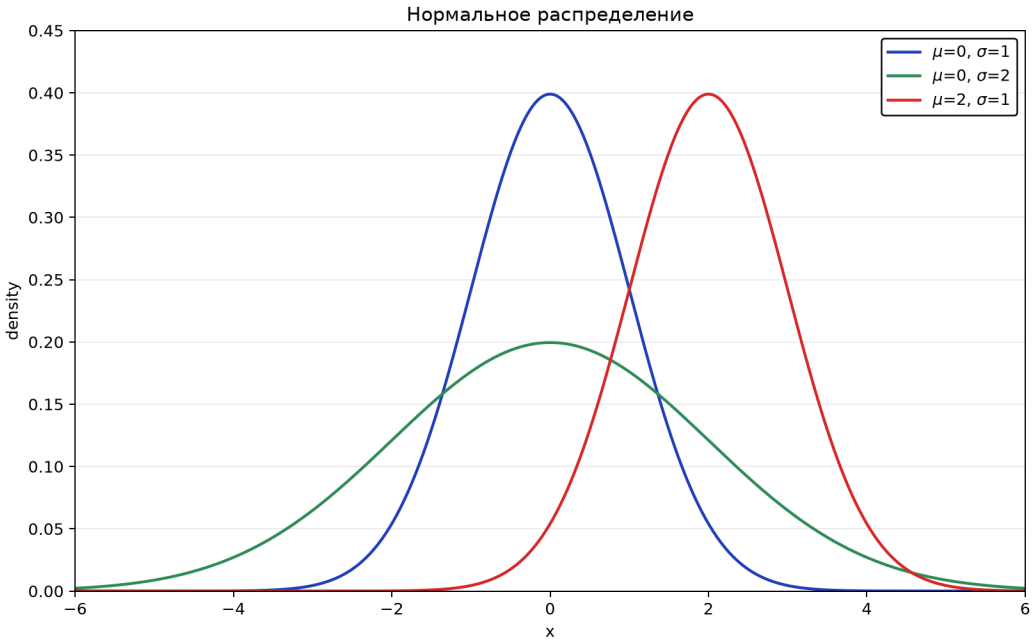


Рис. 2.3. *Плотности нормального распределения при* $(\mu, \sigma) = (0, 1), (0, 2)$ *и* $(2, 1)$

Синие и красные кривые имеют одинаковую ширину $\sigma = 1$ и разные средние: изменение μ сдвигает распределение вдоль оси x. У зелёной кривой $\sigma = 2$, поэтому она шире и ниже. Площадь под каждой кривой по-прежнему равна единице.

2.6.2. Моменты распределения Гаусса

Параметры нормального распределения совпадают с его средним и стандартным отклонением:

$$\mathbb{E}[X] = \mu, \qquad \operatorname{Var}(X) = \sigma^2. \tag{2.26}$$

Вывод

2.6.2.1. Среднее

Перейдём к стандартной нормальной величине

$$z = \frac{x - \mu}{\sigma}, \qquad x = \mu + \sigma z, \qquad dx = \sigma\, dz,$$

плотность которой обозначим

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}. \tag{2.27}$$

Тогда

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{+\infty} (\mu + \sigma z)\varphi(z)\,dz \\ &= \mu \int_{-\infty}^{+\infty} \varphi(z)\,dz + \sigma \int_{-\infty}^{+\infty} z\varphi(z)\,dz \\ &= \mu.\end{aligned} \tag{2.28}$$

Первый интеграл равен единице по условию нормировки. Второй равен нулю, потому что $z\varphi(z)$ — нечётная функция.

Вывод

2.6.2.2. Дисперсия

Та же замена переменной даёт

$$\operatorname{Var}(X) = \sigma^2 \int_{-\infty}^{+\infty} z^2 \varphi(z)\,dz.$$

Оставшийся интеграл вычислим по частям. Поскольку

$$\varphi'(z) = -z\varphi(z),$$

получаем

$$\begin{aligned}\int_{-\infty}^{+\infty} z^2\varphi(z)\,dz &= \left[-z\varphi(z)\right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} \varphi(z)\,dz \\ &= 1.\end{aligned}$$

Следовательно,

$$\operatorname{Var}(X) = \sigma^2. \tag{2.29}$$

2.7. Сравнение трёх распределений

В таблице 2.1 собраны три рассмотренные модели. В последнем столбце перечислены условия, которые относятся к устройству повторяемого эксперимента.

Таблица 2.1. Основные свойства биномиального, пуассоновского и нормального распределений

Распределение	Среднее	Дисперсия	Условия применения
$K \sim \mathrm{Bin}(n, p)$	np	$np(1-p)$	Фиксированное n; независимые испытания; постоянное p
$N \sim \mathrm{Pois}(\lambda)$	λ	λ	Счёт событий в фиксированном интервале; постоянная интенсивность; независимость

Распределение	Среднее	Дисперсия	Условия применения
$X \sim \mathcal{N}(\mu, \sigma^2)$	μ	σ^2	Непрерывная величина; приближение для суммы многих малых независимых вкладов

Итоги главы

- Случайная величина описывается распределением. Для дискретной величины оно задаёт вероятности отдельных значений, для непрерывной — плотность вероятности.
- Математическое ожидание указывает центр распределения, а дисперсия измеряет средний квадрат отклонения от этого центра. Эти величины существуют только при сходимости соответствующих суммы или интеграла.
- Биномиальное распределение относится к фиксированному числу испытаний, распределение Пуассона — к числу событий в фиксированном интервале, а нормальное распределение — к непрерывной величине.
- Выбор распределения включает условия, при которых повторяется эксперимент: независимость испытаний, постоянство вероятности или интенсивности и способ набора данных.

2.8. Задачи

Задача 1. Предел биномиального распределения

Покажите, что при

$$n \to \infty, \qquad p \to 0, \qquad np = \lambda = \text{const}$$

биномиальное распределение переходит в распределение Пуассона:

$$\binom{n}{k} p^k (1-p)^{n-k} \longrightarrow \frac{\lambda^k e^{-\lambda}}{k!}.$$

Подсказка. Положите $p = \lambda/n$ и рассмотрите предел при фиксированном k.

Задача 2. Предел Пуассона к гауссову распределению

Используя формулу Стирлинга, покажите, что при больших λ

$$P(N = n) \approx \frac{1}{\sqrt{2\pi\lambda}} \exp\left[-\frac{(n-\lambda)^2}{2\lambda}\right].$$

Задача 3. Моменты равномерного распределения

Пусть $X \sim U(-1, 1)$. Найдите:

1. $\mathbb{E}[X]$;
2. $\mathbb{E}[X^2]$;
3. $\mathrm{Var}(X)$.

Задача 4. Когда моменты однозначно задают распределение

Пусть

$$\mathbb{E}[X^{2n}] = 1, \qquad \mathbb{E}[X^{2n+1}] = 0.$$

Покажите, что $X = \pm 1$ с равными вероятностями.
Подсказка. Используйте $\mathbb{E}[(X^2 - 1)^2] = 0$.

Задача 5. Восстановление распределения на конечном носителе

Пусть X принимает значения -1, 0, 1, причём

$$\mathbb{E}[X] = 0, \qquad \mathbb{E}[X^2] = \frac{1}{2}.$$

Найдите вероятности трёх значений.

Задача 6. Факториальные моменты распределения Пуассона

Для $N \sim \mathrm{Pois}(\lambda)$ покажите, что

$$\mathbb{E}[N(N-1)\cdots(N-r+1)] = \lambda^r.$$

Литература

[1] Feller, W. *An Introduction to Probability Theory and Its Applications*. John Wiley & Sons. 1971.
[2] Pearson, K. Notes on the History of Correlation. *Biometrika*, **13**, 25–45. 1920. doi:10.1093/biomet/13.1.25.

Глава 3

Характеристические функции и суммы случайных величин

Содержание

Аннотация

В этой главе мы научимся находить распределение суммы независимых случайных величин. Прямой путь приводит к свёртке плотностей. Преобразование Фурье заменяет свёртку произведением и вводит характеристическую функцию. На нескольких знакомых распределениях мы разберём, как она работает, извлечём из неё моменты и кумулянты и подойдём к центральной предельной теореме.

3.1. Распределение суммы независимых случайных величин

Пусть X и Y — независимые случайные величины с плотностями $f_X(x)$ и $f_Y(y)$. Нас интересует сумма

$$S = X + Y. \tag{3.1}$$

Такая задача постоянно возникает в эксперименте. Полное энерговыделение складывается из энерговыделений отдельных частиц, заряд на выходе детектора — из зарядов отдельных фотоэлектронов, а результат измерения — из сигнала и шума.

Значение $S = s$ можно получить из любой пары (x, y), для которой $x + y = s$. После выбора x второе слагаемое равно $s - x$. Независимость позволяет перемножить плотности, а интегрирование собирает вклады от всех возможных x:

$$f_S(s) = \int_{-\infty}^{+\infty} f_X(x) f_Y(s - x)\, dx. \tag{3.2}$$

Интеграл в формуле 3.2 называется **свёрткой**. Для него используется короткая запись

$$f_S = f_X * f_Y. \tag{3.3}$$

Звёздочка здесь обозначает операцию свёртки. Условие независимости существенно: для зависимых величин произведение $f_X(x) f_Y(s - x)$ нужно заменить совместной плотностью $f_{X,Y}(x, s - x)$.

3.1.1. Сумма нескольких величин

Для суммы

$$S_n = X_1 + X_2 + \cdots + X_n \tag{3.4}$$

получается последовательная свёртка:

$$f_{S_n} = f_{X_1} * f_{X_2} * \cdots * f_{X_n}. \tag{3.5}$$

В развёрнутом виде это $(n-1)$-кратный интеграл:

$$\begin{aligned} f_{S_n}(s) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} & f_{X_1}(x_1) \cdots f_{X_{n-1}}(x_{n-1}) \\ & \times f_{X_n}(s - x_1 - \cdots - x_{n-1})\, dx_1 \cdots dx_{n-1}. \end{aligned} \tag{3.6}$$

Каждое новое слагаемое добавляет ещё одно интегрирование. Формула верна и полезна, но с ростом размерности интеграла вычислять его непосредственно хочется всё меньше.

3.2. Пример: распределение Коши

Распределение Коши с параметром положения x_0 и масштабом $\gamma > 0$ имеет плотность

$$f_X(x; x_0, \gamma) = \frac{1}{\pi\gamma\left[1 + \left(\dfrac{x - x_0}{\gamma}\right)^2\right]}. \tag{3.7}$$

Та же форма известна как распределение Лоренца. В физике частиц она связана с нерелятивистской формой линии Брейта—Вигнера, а в спектроскопии описывает лоренцев профиль линии. Релятивистская форма Брейта—Вигнера записывается иначе, поэтому считать эти названия полными синонимами не следует.

При $x_0 = 0$ и $\gamma = 1$ получаем стандартное распределение Коши:

$$f_X(x) = \frac{1}{\pi(1 + x^2)}. \tag{3.8}$$

БИОГРАФИЯ

Огюстен Луи Коши

1789–1857

Французский математик и инженер. Коши внёс определяющий вклад в математический анализ, дифференциальные уравнения и математическую физику. Распределение Коши служит важным примером тяжёлых хвостов: его математическое ожидание и дисперсия не существуют.

Источник: MacTutor; фото MacTutor

Возьмём две независимые величины

$$X, Y \sim \mathrm{Cauchy}(0, 1). \tag{3.9}$$

Для суммы $S = X + Y$ формула свёртки даёт

$$f_S(s) = \frac{1}{\pi^2}\int_{-\infty}^{+\infty} \frac{dx}{(1 + x^2)\left[1 + (s - x)^2\right]}. \tag{3.10}$$

В знаменателе стоит многочлен четвёртой степени по x. Интеграл можно взять разложением на простые дроби или методом вычетов, после чего ответ ещё придётся упростить. Задача поставлена одной строкой, а вычисление уже разрослось. Для суммы многих величин этот путь становится совсем непрактичным.

3.3. Характеристическая функция

Вместо самой плотности рассмотрим её преобразование Фурье. Для случайной величины X **характеристическая функция** определяется как [1]

$$\varphi_X(t) = \mathbb{E}\left[e^{itX}\right]. \tag{3.11}$$

Если X имеет плотность, математическое ожидание в формуле 3.11 записывается интегралом:

$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f_X(x)\, dx. \tag{3.12}$$

Здесь x — переменная интегрирования, а t — аргумент характеристической функции. Само распределение вещественно, но его Фурье-образ в общем случае комплексный.

Характеристическая функция однозначно определяет распределение. Если плотность существует и выполнены обычные условия теоремы обращения Фурье, её можно восстановить по формуле

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \varphi_X(t)\, dt. \tag{3.13}$$

Преобразование Фурье переводит свёртку в произведение. Для независимых X и Y это видно непосредственно:

$$\begin{aligned} \varphi_{X+Y}(t) &= \mathbb{E}\left[e^{it(X+Y)}\right] \\ &= \mathbb{E}\left[e^{itX} e^{itY}\right] \\ &= \mathbb{E}\left[e^{itX}\right] \mathbb{E}\left[e^{itY}\right] \\ &= \varphi_X(t)\varphi_Y(t). \end{aligned} \tag{3.14}$$

Третье равенство и есть место, где используется независимость.

3.4. Сумма двух величин Коши через характеристические функции

Характеристическая функция стандартного распределения Коши равна

$$\varphi_X(t) = e^{-|t|}. \tag{3.15}$$

Её можно получить из формулы 3.8 методом вычетов. Для двух независимых стандартных величин Коши формула 3.14 сразу даёт

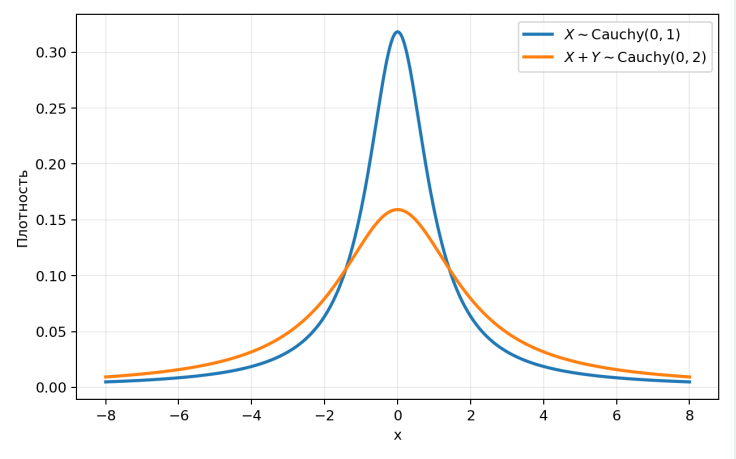
$$\varphi_{X+Y}(t) = e^{-|t|} e^{-|t|} = e^{-2|t|}. \tag{3.16}$$

Последнее выражение является характеристической функцией распределения Коши с масштабом $\gamma = 2$. Поэтому

$$S = X + Y \sim \mathrm{Cauchy}(0, 2), \qquad f_S(s) = \frac{2}{\pi(s^2 + 4)}. \tag{3.17}$$

Мы получили ответ, не вычисляя интеграл 3.10.

Стандартное распределение Коши и распределение суммы двух независимых стандартных величин Коши.



Интерактив. Сумма двух величин Коши
В HTML-книге график строится из характеристических функций исходной величины и её суммы с независимой копией.

У суммы максимум ниже, а распределение шире: параметр γ увеличился с 1 до 2.

3.4.1. Вычисление распределения суммы через Фурье

Весь метод состоит из трёх операций.

01 Преобразовать плотности

Найти характеристические функции

$$\varphi_X(t) = \int e^{itx} f_X(x)\,dx, \qquad \varphi_Y(t) = \int e^{ity} f_Y(y)\,dy. \tag{3.18}$$

↓

02 Перемножить

Для независимой суммы использовать

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t). \tag{3.19}$$

↓

03 Вернуться к плотности

Выполнить обратное преобразование

$$f_{X+Y}(s) = \frac{1}{2\pi}\int e^{-its}\varphi_{X+Y}(t)\,dt. \tag{3.20}$$

Для десяти слагаемых вместо девятикратной свёртки получается произведение десяти функций и одно обратное преобразование.

Тот же приём используется при восстановлении сигнала на фоне независимого шума. Если $D = S + N$, то $\varphi_D = \varphi_S \varphi_N$. Измерив шум отдельно, можно найти φ_S делением. На практике это деление усиливает высокочастотные флуктуации и требует регуляризации. Преобразование Фурье упрощает алгебру, но не улучшает качество данных.

3.5. Свойства характеристической функции

3.5.1. Дискретный случай

Если X принимает значения x_k с вероятностями p_k, интеграл заменяется суммой:

$$\varphi_X(t) = \sum_k p_k e^{itx_k}. \tag{3.21}$$

Это то же распределение, записанное в форме, удобной для сложения случайных величин.

3.5.2. Почему характеристическая функция всегда существует

Модуль комплексной экспоненты равен единице. Поэтому

$$|\varphi_X(t)| = \left|\mathbb{E}[e^{itX}]\right| \le \mathbb{E}[|e^{itX}|] = 1. \tag{3.22}$$

В отличие от математического ожидания $\mathbb{E}[X]$ или дисперсии, интеграл в определении характеристической функции не страдает от тяжёлых хвостов: подынтегральная функция по модулю ограничена единицей. Характеристическая функция существует у любого распределения.

3.5.3. Базовые свойства

Для любой случайной величины выполняются соотношения

$$\varphi_X(0) = 1, \qquad |\varphi_X(t)| \le 1, \qquad \varphi_X(-t) = \varphi_X^*(t). \tag{3.23}$$

Первое равенство выражает нормировку вероятности: $\varphi_X(0) = \mathbb{E}[1] = 1$. Третье следует из комплексного сопряжения экспоненты. Для распределения, симметричного относительно нуля, характеристическая функция вещественна и чётна.

3.5.4. Сдвиг и масштаб

Пусть $Y = aX + b$. Тогда

$$\begin{aligned} \varphi_Y(t) &= \mathbb{E}[e^{it(aX+b)}] \\ &= e^{itb}\mathbb{E}[e^{i(at)X}] \\ &= e^{itb}\varphi_X(at). \end{aligned} \tag{3.24}$$

Сдвиг случайной величины даёт фазовый множитель, а изменение масштаба меняет аргумент характеристической функции.

3.6. Суммы многих независимых величин

Последовательно применяя формулу 3.14 к $S_n = X_1 + \cdots + X_n$, получаем

$$\varphi_{S_n}(t) = \prod_{k=1}^{n} \varphi_{X_k}(t). \tag{3.25}$$

Если все слагаемые имеют одно распределение,

$$\varphi_{S_n}(t) = [\varphi_X(t)]^n. \tag{3.26}$$

Ради этих двух формул характеристические функции и вводят в задачах о суммах.

3.7. Примеры характеристических функций

3.7.1. Распределение Бернулли

Величина Бернулли принимает значение 1 с вероятностью p и значение 0 с вероятностью $1-p$. По формуле 3.21

$$\varphi_X(t) = (1-p)e^{it\cdot 0} + pe^{it\cdot 1} = (1-p) + pe^{it}. \tag{3.27}$$

Сумма $S_n = X_1 + \cdots + X_n$ считает число единиц в n независимых испытаниях. Её характеристическая функция равна

$$\begin{aligned} \varphi_{S_n}(t) &= \left[(1-p) + pe^{it}\right]^n \\ &= \sum_{k=0}^{n} \binom{n}{k} (1-p)^{n-k} p^k e^{itk}. \end{aligned} \tag{3.28}$$

Коэффициент при e^{itk} равен вероятности $P(S_n = k)$. Мы снова получили биномиальное распределение:

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad S_n \sim \mathrm{Bin}(n, p). \tag{3.29}$$

Другого ответа и быть не могло: сумма нулей и единиц как раз считает число успехов. Характеристическая функция показывает это непосредственно в алгебраической форме.

3.7.2. Распределение Пуассона

Для $N \sim \mathrm{Pois}(\lambda)$

$$\begin{aligned} \varphi_N(t) &= \sum_{k=0}^{\infty} e^{itk} e^{-\lambda} \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} \\ &= \exp\left[\lambda(e^{it} - 1)\right]. \end{aligned} \tag{3.30}$$

Пусть независимые величины N_k имеют параметры λ_k. Тогда

$$\begin{aligned} \varphi_{\sum_k N_k}(t) &= \prod_k \exp\left[\lambda_k(e^{it} - 1)\right] \\ &= \exp\left[\left(\sum_k \lambda_k\right)(e^{it} - 1)\right]. \end{aligned} \tag{3.31}$$

Получилась характеристическая функция Пуассона с параметром $\sum_k \lambda_k$:

$$\sum_k N_k \sim \mathrm{Pois}\left(\sum_k \lambda_k\right). \tag{3.32}$$

Это соответствует устройству опыта. Если несколько независимых источников дают в среднем $\lambda_1, \lambda_2, \ldots$ событий за один интервал времени, общий счёт имеет среднее $\lambda_1 + \lambda_2 + \cdots$.

3.7.3. Распределение Гаусса

Для $X \sim \mathcal{N}(\mu, \sigma^2)$ характеристическая функция имеет вид

$$\varphi_X(t) = \exp\left(i\mu t - \frac{\sigma^2 t^2}{2}\right). \tag{3.33}$$

Вывод

3.7.3.1. Вывод характеристической функции Гаусса

Сначала рассмотрим стандартную нормальную величину $Z \sim \mathcal{N}(0,1)$:

$$\varphi_Z(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{itz} e^{-z^2/2}\, dz. \tag{3.34}$$

Продифференцируем по t и проинтегрируем по частям, используя $ze^{-z^2/2} = -d(e^{-z^2/2})/dz$:

$$\begin{aligned} \varphi_Z'(t) &= \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} ze^{itz} e^{-z^2/2}\, dz \\ &= -t\varphi_Z(t). \end{aligned} \tag{3.35}$$

Условие $\varphi_Z(0) = 1$ выбирает решение

$$\varphi_Z(t) = e^{-t^2/2}. \tag{3.36}$$

Поскольку $X = \mu + \sigma Z$, правило сдвига и масштаба 3.24 даёт формулу 3.33.

Для независимых $X_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$ произведение характеристических функций равно

$$\varphi_{\sum_k X_k}(t) = \exp\left[it \sum_k \mu_k - \frac{t^2}{2} \sum_k \sigma_k^2\right]. \tag{3.37}$$

Это снова гауссова характеристическая функция. Следовательно,

$$\sum_k X_k \sim \mathcal{N}\left(\sum_k \mu_k, \sum_k \sigma_k^2\right). \tag{3.38}$$

У независимых гауссовых величин складываются средние и дисперсии. Стандартные отклонения складывать нельзя.

3.7.4. Распределение Коши

Для независимых $X_k \sim \mathrm{Cauchy}(0,1)$

$$\varphi_{S_n}(t) = \left(e^{-|t|}\right)^n = e^{-n|t|}, \qquad S_n = \sum_{k=1}^{n} X_k. \tag{3.39}$$

Поэтому $S_n \sim \text{Cauchy}(0, n)$. После деления суммы на n масштаб возвращается к единице:

$$\frac{S_n}{n} \sim \text{Cauchy}(0, 1). \tag{3.40}$$

Среднее арифметическое имеет то же распределение, что и каждое слагаемое. Оно не сжимается с ростом n. К этому примеру мы вернёмся при обсуждении условий центральной предельной теоремы.

3.8. Моменты из характеристической функции

Разложим комплексную экспоненту около $t = 0$:

$$e^{itX} = \sum_{n=0}^{\infty} \frac{(itX)^n}{n!}. \tag{3.41}$$

Если соответствующие моменты существуют и разрешено переставить математическое ожидание и сумму, то

$$\varphi_X(t) = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \mathbb{E}[X^n]. \tag{3.42}$$

Отсюда моменты извлекаются производными в нуле:

$$\mathbb{E}[X^n] = \frac{1}{i^n} \varphi_X^{(n)}(0). \tag{3.43}$$

В частности,

$$\mathbb{E}[X] = \frac{\varphi_X'(0)}{i}, \qquad \mathbb{E}[X^2] = -\varphi_X''(0). \tag{3.44}$$

Слово «если» перед формулой 3.42 нельзя выбрасывать. Сама характеристическая функция существует всегда, её производные нужного порядка — нет.

Для стандартного распределения Коши

$$\varphi_X(t) = e^{-|t|}. \tag{3.45}$$

У этой функции нет производной при $t = 0$. Тот же факт виден в обычном пространстве: интеграл для $\mathbb{E}[X]$ не сходится абсолютно. Односторонняя производная или главное значение интеграла не заменяют математического ожидания.

3.9. Кумулянты

Произведение характеристических функций удобно превратить в сумму. Для этого берётся логарифм:

$$K_X(t) = \ln \varphi_X(t). \tag{3.46}$$

Поскольку $\varphi_X(0) = 1$ и φ_X непрерывна, в некоторой окрестности нуля она не обращается в нуль. Там можно выбрать непрерывную ветвь логарифма с $K_X(0) = 0$. Если нужные производные существуют, разложение имеет вид

$$K_X(t) = \sum_{n=1}^{\infty} \kappa_n \frac{(it)^n}{n!}, \tag{3.47}$$

а коэффициенты κ_n называются **кумулянтами**:

$$\kappa_n = \frac{1}{i^n}\,\frac{d^n}{dt^n}\ln \varphi_X(t)\Big|_{t=0}. \tag{3.48}$$

3.9.1. Первые четыре кумулянта

Для первой производной

$$K_X'(0) = \frac{\varphi_X'(0)}{\varphi_X(0)} = i\mathbb{E}[X]. \tag{3.49}$$

Для второй

$$\begin{aligned} K_X''(0) &= \varphi_X''(0) - \left[\varphi_X'(0)\right]^2 \\ &= -\mathbb{E}[X^2] + \mathbb{E}[X]^2 \\ &= -\operatorname{Var}(X). \end{aligned} \tag{3.50}$$

После деления на соответствующие степени i получаем

$$\begin{aligned} \kappa_1 &= \mu, \\ \kappa_2 &= \sigma^2, \\ \kappa_3 &= \mathbb{E}[(X-\mu)^3], \\ \kappa_4 &= \mathbb{E}[(X-\mu)^4] - 3\sigma^4. \end{aligned} \tag{3.51}$$

Первые два кумулянта — среднее и дисперсия. Третий равен третьему центральному моменту; после деления на σ^3 он даёт коэффициент асимметрии. Четвёртый измеряет отклонение четвёртого центрального момента от гауссова значения $3\sigma^4$; деление на σ^4 даёт эксцесс.

3.9.2. Почему кумулянты удобны для сумм

Для независимых X и Y

$$\begin{aligned} K_{X+Y}(t) &= \ln[\varphi_X(t)\varphi_Y(t)] \\ &= K_X(t) + K_Y(t). \end{aligned} \tag{3.52}$$

Значит, каждый существующий кумулянт суммы равен сумме кумулянтов:

$$\kappa_n(X+Y) = \kappa_n(X) + \kappa_n(Y). \tag{3.53}$$

Для изменения масштаба действует правило

$$\kappa_n(aX) = a^n \kappa_n(X). \tag{3.54}$$

У обычных моментов суммы появляются смешанные члены. Кумулянты независимых слагаемых складываются без таких членов, что особенно удобно для нормированных сумм.

3.10. Мост к центральной предельной теореме

Пусть $X_1, \dots, X_n$ независимы, имеют одно распределение и конечные среднее и дисперсию:

$$\mathbb{E}[X_k] = \mu, \qquad \operatorname{Var}(X_k) = \sigma^2 < \infty. \tag{3.55}$$

Рассмотрим центрированную и нормированную сумму

$$Z_n = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} (X_k - \mu). \tag{3.56}$$

Введём $Y_k = (X_k - \mu)/\sigma$. Тогда $\mathbb{E}[Y_k] = 0$, $\mathrm{Var}(Y_k) = 1$ и

$$\varphi_{Z_n}(t) = \left[\varphi_Y\!\left(\frac{t}{\sqrt{n}} \right) \right]^n. \tag{3.57}$$

Конечность второго момента даёт при $u \to 0$ разложение

$$\varphi_Y(u) = 1 - \frac{u^2}{2} + o(u^2). \tag{3.58}$$

Подставляя $u = t/\sqrt{n}$, получаем

$$\varphi_{Z_n}(t) = \left[1 - \frac{t^2}{2n} + o\!\left(\frac{1}{n}\right) \right]^n. \tag{3.59}$$

При каждом фиксированном t

$$\lim_{n\to\infty} \varphi_{Z_n}(t) = e^{-t^2/2}. \tag{3.60}$$

Справа стоит характеристическая функция $\mathcal{N}(0,1)$. Теорема непрерывности Леви переводит сходимость характеристических функций в сходимость по распределению [1]. Полную формулировку центральной предельной теоремы и скорость приближения мы разберём в следующей главе.

Нормировка в формуле 3.56 обязательна. Сама сумма имеет среднее $n\mu$ и дисперсию $n\sigma^2$; её центр и ширина растут вместе с n. Для среднего арифметического та же нормированная величина записывается как

$$Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}. \tag{3.61}$$

Отсюда возникает знакомый масштаб статистической неопределённости среднего $1/\sqrt{n}$.

3.10.1. Тот же предел через кумулянты

Если кумулянт порядка r существует, правила 3.53 и 3.54 дают

$$\kappa_r(Z_n) = n^{1-r/2}\kappa_r(Y). \tag{3.62}$$

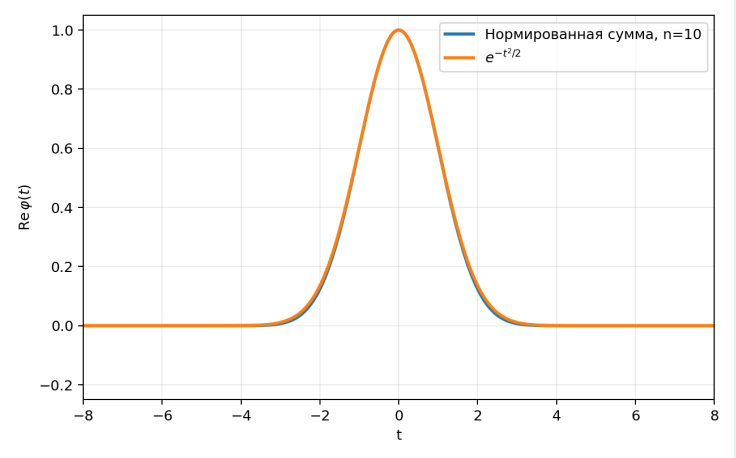
При $r = 2$ дисперсия остаётся равной единице. Третий кумулянт убывает как $1/\sqrt{n}$, четвёртый — как $1/n$, а каждый существующий кумулянт порядка $r > 2$ стремится к нулю. Если высшие кумулянты существуют, в разложении остаётся квадратичный член $\ln\varphi(t) = -t^2/2$, соответствующий распределению Гаусса.

3.11. Интерактив: приближение к распределению Гаусса в пространстве Фурье

Возьмём независимые величины

$$X_k \sim U(-\sqrt{3}, \sqrt{3}), \qquad \mathbb{E}[X_k] = 0, \qquad \mathrm{Var}(X_k) = 1. \tag{3.63}$$

На графике сравниваются характеристическая функция нормированной суммы и предельная функция $e^{-t^2/2}$. Ползунок меняет число слагаемых n.



Интерактив. Гауссов предел в пространстве Фурье
Меняя число слагаемых, можно проследить, как характеристическая функция нормированной суммы приближается к $e^{-t^2/2}$.

Итоги главы

- Характеристическая функция является преобразованием Фурье распределения и существует для любой случайной величины.
- Для суммы независимых величин характеристические функции перемножаются. Это заменяет последовательные свёртки плотностей обычным произведением.
- Моменты связаны с производными $\varphi_X(t)$ в нуле, а кумулянты — с производными $\ln \varphi_X(t)$. Характеристическая функция существует всегда; нужных производных у неё может не быть.
- У нормированной суммы сохраняется дисперсия, а вклад высших существующих кумулянтов убывает с ростом числа слагаемых. Так характеристические функции приводят к гауссову пределу.

3.12. Задачи

Задача 1. Сумма двух равномерных величин

Пусть независимые случайные величины распределены равномерно:

$$X, Y \sim \mathrm{U}(0,1), \qquad f_X(x) = f_Y(x) = \theta(x)\theta(1-x).$$

Для суммы $S = X + Y$:

1. Вычислите плотность $f_S(s)$ непосредственно как свёртку.
2. Покажите, что произведение θ-функций ограничивает область интегрирования условием
$$\max(0, s-1) \le x \le \min(1, s).$$
3. Получите кусочно заданную треугольную плотность и проверьте её нормировку.
4. Найдите характеристическую функцию $\varphi_X(t)$.
5. Используя $\varphi_S(t) = \varphi_X(t)^2$, получите тот же ответ через обратное преобразование Фурье.
6. Нарисуйте плотности $f_X(x)$ и $f_S(s)$ и объясните форму распределения суммы.

Задача 2. Распределение Бернулли

Пусть
$$P(X = 1) = p, \qquad P(X = 0) = 1 - p.$$

1. Найдите $\varphi_X(t)$.
2. Найдите характеристическую функцию суммы n независимых величин Бернулли.
3. Объясните, почему сумма имеет биномиальное распределение.

Задача 3. Сумма пуассоновских величин

Для независимых
$$N_1 \sim \mathrm{Pois}(\lambda_1), \qquad N_2 \sim \mathrm{Pois}(\lambda_2)$$
покажите с помощью характеристических функций, что
$$N_1 + N_2 \sim \mathrm{Pois}(\lambda_1 + \lambda_2).$$

Задача 4. Сумма гауссовых величин

Пусть независимые величины распределены как
$$X_k \sim \mathrm{N}(\mu_k, \sigma_k^2).$$
Покажите, что
$$\sum_k X_k \sim \mathrm{N}\left(\sum_k \mu_k, \sum_k \sigma_k^2\right).$$

Задача 5. Моменты и производные

Для

$$\varphi_N(t) = \exp\bigl[\lambda(e^{it} - 1)\bigr]$$

найдите $\mathbb{E}[N]$, $\mathbb{E}[N^2]$ и $\mathrm{Var}(N)$ через производные в нуле.

Задача 6. Распределение Коши

Для

$$\varphi_X(t) = e^{-|t|}$$

объясните, почему $\varphi_X'(0)$ не существует, и покажите, что

$$\frac{X_1 + \cdots + X_n}{n}$$

имеет то же распределение Коши.

Задача 7. Распределение $S_n = X_1 + \cdots + X_n$ в пределе $n \to \infty$

Пусть $X_1, \ldots, X_n$ — независимые одинаково распределённые случайные величины,

$$S_n = X_1 + \cdots + X_n, \qquad \langle X \rangle_n = \frac{S_n}{n}.$$

Среднее и вариация X_i:

$$\mathbb{E}[X_i] = \mu, \qquad \mathrm{Var}(X_i) = \sigma^2 < \infty.$$

Рассмотрим нормированную сумму

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

- Покажите, что характеристическая функция Z_n имеет вид

$$\varphi_{Z_n}(t) = \left[e^{-it\mu/(\sigma\sqrt{n})} \varphi_X\left(\frac{t}{\sigma\sqrt{n}}\right) \right]^n.$$

- Для каждого из следующих распределений найдите характеристическую функцию и предел

$$\lim_{n\to\infty} \varphi_{Z_n}(t).$$

Симметричное распределение Бернулли:

$$P(X = 1) = P(X = -1) = \frac{1}{2}.$$

Равномерное распределение:

$$X \sim U(-\sqrt{3}, \sqrt{3}).$$

Экспоненциальное распределение:

$$X \sim \mathrm{Exp}(1).$$

Распределение Пуассона:

$$X \sim \mathrm{Pois}(\lambda).$$

- Убедитесь, что во всех четырёх случаях получается один и тот же предел:

$$\varphi_{Z_n}(t) \to e^{-t^2/2}.$$

Литература

[1] Feller, W. *An Introduction to Probability Theory and Its Applications.* John Wiley & Sons. 1971.

Глава 4

Центральная предельная теорема и нормальное распределение

Содержание

Аннотация

Центральная предельная теорема - одна из важнейших теорем в статистическом анализе. Она показывает фундаментально важную роль нормального распределения. Мы увидим как нормальное или гауссовое распределение возникает почти волшебным образом для описания суммы бесконечного количества случайных величин. Причём каждая из них может следовать почти любому распределению. Мы выведем среднее и дисперсию суммы, разберём точную формулировку теоремы и проследим гауссов предел в численном эксперименте. Нормальное распределение затем станет рабочим инструментом: через его квантили мы построим центральные области и сравним два независимых измерения.

4.1. Физическая задача: сигнал фотодетектора

Пусть короткая вспышка света выбивает в фотодетекторе N фотоэлектронов. Каждый фотоэлектрон создаёт на выходе электрический импульс со случайным зарядом Q_i. Электроника измеряет их сумму

$$Q_N = Q_1 + Q_2 + \cdots + Q_N. \tag{4.1}$$

Даже при одинаковых вспышках заряды отдельных импульсов немного различаются. Обозначим средний заряд одного импульса через q_0, а его дисперсию — через s_q^2:

$$\mathbb{E}[Q_i] = q_0, \qquad \mathrm{Var}(Q_i) = s_q^2. \tag{4.2}$$

В простейшей модели импульсы независимы и подчиняются одному распределению. Тогда

$$\mathbb{E}[Q_N] = Nq_0, \qquad \mathrm{Var}(Q_N) = Ns_q^2. \tag{4.3}$$

Средний сигнал растёт как N, его стандартное отклонение — как $\sqrt{N}$. Относительная ширина пика равна

$$\frac{\sigma_{Q_N}}{\mathbb{E}[Q_N]} = \frac{s_q}{q_0\sqrt{N}}. \tag{4.4}$$

Если среднее число фотоэлектронов пропорционально энергии, этот результат даёт знакомый стохастический вклад в энергетическое разрешение, пропорциональный $1/\sqrt{E}$. Остаётся понять форму пика. Центральная предельная теорема даёт ответ:

$$\frac{Q_N - Nq_0}{s_q\sqrt{N}} \xrightarrow[N\to\infty]{d} \mathcal{N}(0,1). \tag{4.5}$$

Вся задача уложилась в среднее суммы, дисперсию суммы и одну предельную теорему. В реальном детекторе флуктуирует также число фотоэлектронов, бывают корреляции, неоднородности и шум электроники. Они меняют коэффициент и добавляют другие слагаемые к разрешению. Упрощённая модель показывает сам механизм: средние независимых вкладов складываются, и их дисперсии тоже складываются.

Теперь разберём этот расчёт по частям.

4.2. Псевдослучайные числа

Для проверки центральной предельной теоремы устроим численный эксперимент. В первой главе мы уже генерировали псевдоданные и подбирали ширину распределения. Откуда компьютер берёт случайные числа?

Он этого не умеет делать! Любой генератор случайных чисел генерирует последовательность псевдослучайных чисел, потому что его программа представляет собой детерминированный алгоритм. Внутреннее состояние генератора полностью определяет следующее число. При удачно выбранном алгоритме последовательность проходит статистические тесты и имеет период, намного превышающий длину обычного расчёта. Для моделирования этого достаточно, хотя физического источника случайности внутри алгоритма нет.

Начальное состояние генератора задаётся некоторым *начальным* числом, по английски *seed*. Одинаковое начальное состояние воспроизводит ту же последовательность. Это удобно при проверке программы: рисунок или численный результат можно получить ещё раз.

Генератор равномерного распределения на $[-1, 1]$ в NumPy вызывается несколькими строками:

```
import numpy as np

rng = np.random.default_rng(7)
x = rng.uniform(-1, 1, size=20_000)
```

На рисунке 4.1 показана гистограмма такой выборки. Отдельные числа ничего не сообщают о форме распределения. На большой выборке становится видна постоянная плотность $f(x) = 1/2$ внутри заданного интервала.

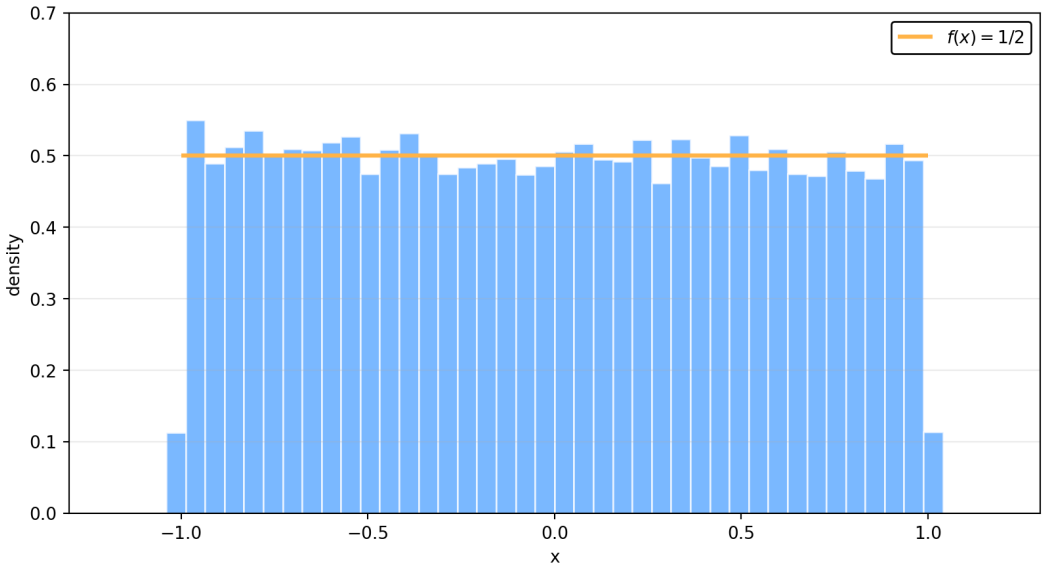


Рис. 4.1. *Гистограмма 20 000 псевдослучайных чисел, равномерно распределённых на* $[-1, 1]$*; горизонтальная линия показывает плотность* $f(x) = 1/2$

Гистограмма сама остаётся случайной: при другом начальном состоянии высоты столбцов изменятся. Распределение и его параметры при этом заданы в программе и остаются теми же.

4.3. Среднее и дисперсия суммы

Рассмотрим сумму m случайных величин

$$S_m = X_1 + X_2 + \cdots + X_m. \tag{4.6}$$

Математическое ожидание линейно:

$$\mathbb{E}[S_m] = \sum_{i=1}^{m} \mathbb{E}[X_i]. \tag{4.7}$$

Для этого равенства независимость не требуется. Со средним всё складывается буквально. При вычислении дисперсии появляются произведения разных слагаемых. Их описывает ковариация.

Ковариация случайных величин X и Y определяется равенством

$$\begin{aligned} \operatorname{cov}(X,Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \end{aligned} \tag{4.8}$$

При $X = Y$ ковариация превращается в дисперсию: $\operatorname{cov}(X, X) = \operatorname{Var}(X)$.

Вывод

4.3.1. Дисперсия суммы

Вычтем из каждого слагаемого его среднее и раскроем квадрат:

$$\begin{aligned} \operatorname{Var}(S_m) &= \mathbb{E}\left[\left(\sum_{i=1}^{m}(X_i - \mathbb{E}[X_i])\right)^2\right] \\ &= \sum_{i=1}^{m} \operatorname{Var}(X_i) + 2\sum_{i<j} \operatorname{cov}(X_i, X_j). \end{aligned} \tag{4.9}$$

Для независимых величин $\mathbb{E}[X_i X_j] = \mathbb{E}[X_i]\mathbb{E}[X_j]$ при $i \neq j$, поэтому все ковариации в двойной сумме равны нулю:

$$\operatorname{Var}(S_m) = \sum_{i=1}^{m} \operatorname{Var}(X_i). \tag{4.10}$$

Результат из уравнения 4.10 мы уже использовали для фотодетектора. Стандартные отклонения независимых вкладов складываются в квадратуре. Если между вкладами есть корреляции, двойная сумма в уравнении 4.9 сохраняется.

4.3.2. Равномерные слагаемые

Для численного эксперимента возьмём независимые одинаково распределённые величины

$$X_i \sim U(-1, 1). \tag{4.11}$$

Симметрия даёт $\mathbb{E}[X_i] = 0$. Плотность равна $1/2$ на $[-1, 1]$, поэтому

$$\begin{aligned} \mathbb{E}[X_i^2] &= \frac{1}{2}\int_{-1}^{1} x^2\,dx = \frac{1}{3}, \\ \operatorname{Var}(X_i) &= \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = \frac{1}{3}. \end{aligned} \tag{4.12}$$

Для суммы S_m получаем

$$\mathbb{E}[S_m] = 0, \qquad \operatorname{Var}(S_m) = \frac{m}{3}. \tag{4.13}$$

Ширина самой суммы растёт с m. Чтобы сравнивать её форму при разных m, нужно каждый раз вернуть среднее к нулю, а дисперсию — к единице:

$$Z_m = \frac{S_m}{\sqrt{m/3}} = \frac{X_1 + \cdots + X_m}{\sqrt{m/3}}. \tag{4.14}$$

4.4. Центральная предельная теорема

Теперь можно сформулировать теорему без сокращений.

Центральная предельная теорема. Пусть $X_1, \ldots, X_m$ независимы, одинаково распределены и имеют конечные среднее и ненулевую дисперсию:

$$\mathrm{E}[X_i] = \mu, \qquad 0 < \mathrm{Var}(X_i) = \sigma^2 < \infty.$$

Тогда

$$Z_m = \frac{\sum_{i=1}^{m} X_i - m\mu}{\sigma\sqrt{m}} \xrightarrow[m\to\infty]{d} \mathcal{N}(0,1). \tag{4.15}$$

Стрелка $\xrightarrow{d}$ означает сходимость по распределению. При каждом фиксированном z в точке непрерывности предельной функции распределения

$$P(Z_m \le z) \longrightarrow \Phi(z),$$

где Φ — CDF стандартного нормального распределения. В предыдущей главе мы увидели тот же предел через характеристические функции.

Для независимых, но неодинаково распределённых слагаемых одной конечности дисперсий мало. Один член суммы может сохранить заметную долю полного разброса и определять форму результата. Условия Линдеберга и Ляпунова исключают такое доминирование; соответствующие варианты теоремы приведены, например, в [1].

БИОГРАФИЯ

Александр Михайлович Ляпунов

1857–1918

Русский математик, известный работами по устойчивости движения и теории вероятностей. В начале XX века Ляпунов доказал вариант центральной предельной теоремы методом характеристических функций. Условие, названное его именем, ограничивает вклад далёких значений отдельных слагаемых.

Источник: MacTutor; фото MacTutor

В формуле 4.15 существенны центрирование и нормировка. Среднее ненормированной суммы равно $m\mu$, а стандартное отклонение — $\sigma\sqrt{m}$. И центр, и ширина уходят с ростом m. Для выборочного среднего $\overline{X}_m = S_m/m$ та же нормированная величина имеет вид

$$Z_m = \frac{\overline{X}_m - \mu}{\sigma/\sqrt{m}}. \tag{4.16}$$

Отсюда появляется масштаб статистической неопределённости среднего $\sigma/\sqrt{m}$.

Теорема описывает распределение результатов повторяемого опыта. Она ничего не обещает отдельной реализации Z_m: значения порядка двух или трёх стандартных отклонений сохраняют ненулевую вероятность и при сколь угодно большом m.

Универсального числа слагаемых, после которого разрешается написать «распределение уже гауссово», тоже нет. Если существует третий абсолютный центральный момент

$$\rho_3 = \mathbb{E}\left[|X_i - \mu|^3\right],$$

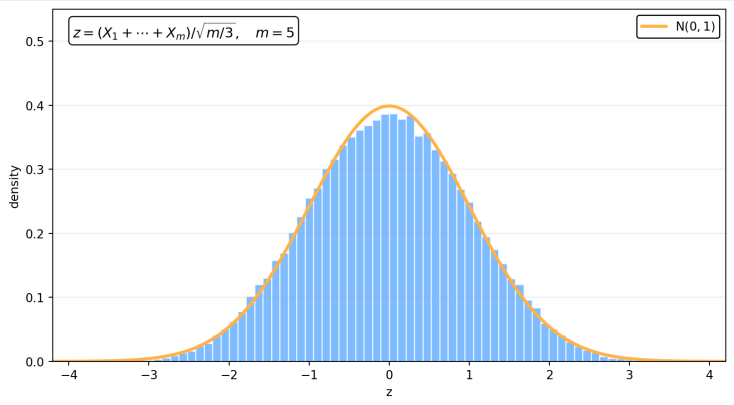
неравенство Берри—Эссеена даёт оценку

$$\sup_z |P(Z_m \le z) - \Phi(z)| \le C\frac{\rho_3}{\sigma^3\sqrt{m}}, \tag{4.17}$$

где C — универсальная константа. Отношение ρ_3/σ^3 зависит от исходного распределения. Кроме того, малая абсолютная ошибка CDF может быть заметной относительной ошибкой в далёком хвосте. При поиске редких сигналов проверка гауссова приближения в центральной области ещё не решает задачу для хвостовой вероятности.

4.4.1. Как меняется форма суммы

Для равномерных слагаемых из уравнения 4.11 при $m = 1$ распределение остаётся равномерным. При $m = 2$ свёртка двух прямоугольников даёт треугольник. Дальше углы сглаживаются, и распределение нормированной суммы приближается к стандартному гауссу.



Интерактив. Форма нормированной суммы
Меняйте число слагаемых и объём выборки. Гистограмма показывает новую реализацию численного эксперимента, а гладкая кривая — плотность $\mathcal{N}(0,1)$.

Вернёмся к фотодетектору. Формула 4.5 является прямым применением ЦПТ к зарядам однофотоэлектронных импульсов. Гауссова форма пика и закон $1/\sqrt{N}$ имеют одно происхождение. Первое относится к форме распределения нормированной суммы, второе — к её относительной ширине.

4.5. Нормальное распределение

Нормальное, или гауссово, распределение уже появлялось в главах 2 и 3. Здесь оно нужно как числовая шкала отклонений.

Случайная величина $X \sim \mathcal{N}(\mu, \sigma^2)$ имеет плотность

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \qquad \sigma > 0. \tag{4.18}$$

Плотность достигает максимума при $x = \mu$, математическое ожидание равно μ, а дисперсия — σ^2. В записи $\mathcal{N}(\mu, \sigma^2)$ вторым параметром стоит именно дисперсия.

На рисунке 4.2 показана плотность при $\mu = 0$ и $\sigma = 1$.

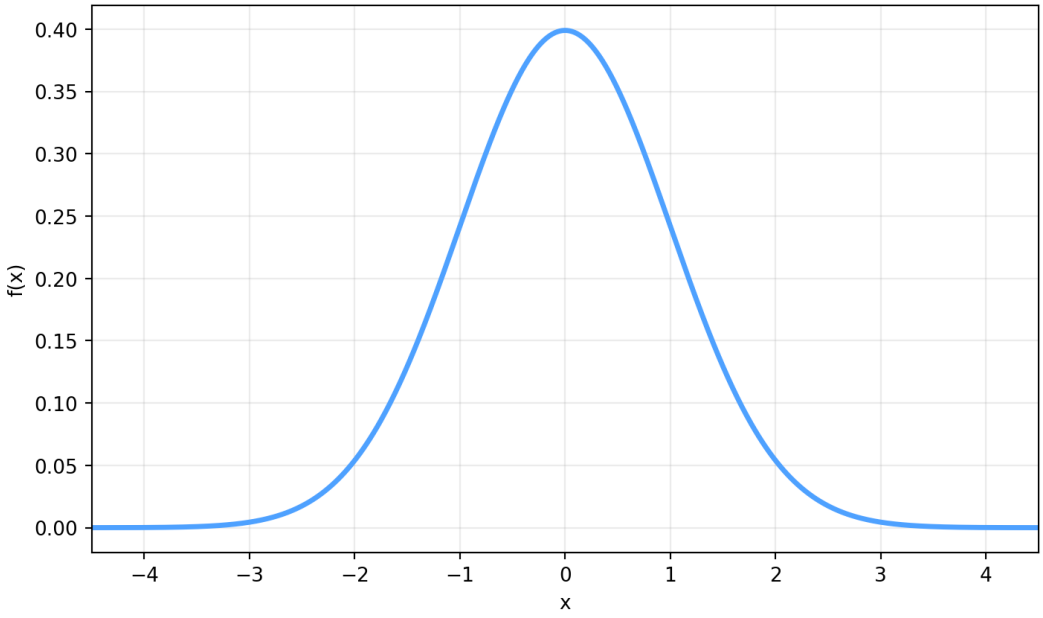


Рис. 4.2. *Плотность стандартного нормального распределения с математическим ожиданием* 0 *и дисперсией* 1

Площадь под всей кривой равна единице. Высота кривой является плотностью; вероятность задаётся площадью над выбранным интервалом оси x.

4.5.1. Стандартизация

Если $X \sim \mathcal{N}(\mu, \sigma^2)$, то линейное преобразование

$$Z = \frac{X - \mu}{\sigma} \tag{4.19}$$

даёт $Z \sim \mathcal{N}(0, 1)$. Число z показывает отклонение от среднего в единицах стандартного отклонения. Результат $x = \mu + 2\sigma$ соответствует $z = 2$ при любых μ и σ.

4.6. Квантили стандартного нормального распределения

Кумулятивная функция стандартной нормальной величины равна

$$\begin{aligned} \Phi(z) &= P(Z \le z) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt \\ &= \frac{1}{2}\left[1 + \operatorname{erf}\!\left(\frac{z}{\sqrt{2}}\right)\right]. \end{aligned} \tag{4.20}$$

Квантиль уровня β — это число z_β, для которого

$$\Phi(z_\beta) = \beta. \tag{4.21}$$

Иными словами, слева от z_β лежит доля вероятности β. Несколько часто используемых значений собраны в таблице 4.1.

Таблица 4.1. Квантили стандартного нормального распределения

β	z_β
0.900	1.282
0.950	1.645
0.975	1.960
0.995	2.576

Рисунок 4.3 показывает, как горизонтальный уровень β на графике CDF определяет квантиль z_β.

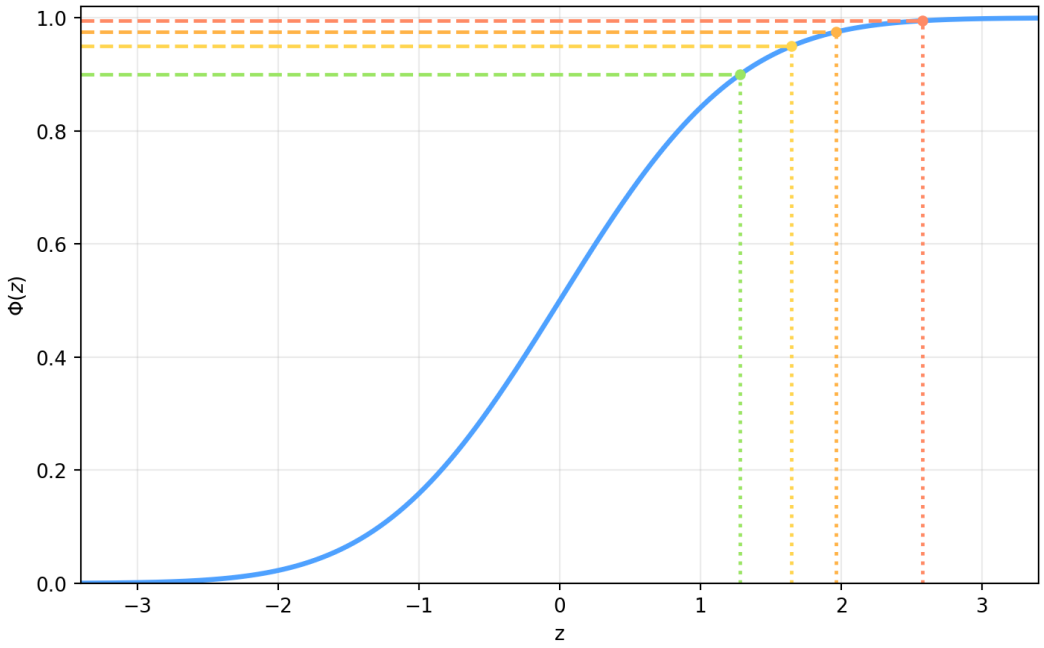


Рис. 4.3. *Кумулятивная функция стандартного нормального распределения; пунктирные линии отмечают уровни β и соответствующие квантили z_β*

4.7. Центральные интервалы

Квантиль z_β из уравнения 4.21 отсекает один левый хвост. Для симметричного распределения часто выбирают центральную область и оставляют в обоих хвостах одинаковую вероятность.

Пусть вероятность центральной области равна γ, а полная вероятность двух хвостов равна $\alpha = 1 - \gamma$. Тогда

$$P(-c \le Z \le c) = \gamma, \qquad P(Z < -c) = P(Z > c) = \frac{\alpha}{2}, \tag{4.22}$$

и граница области выражается через квантиль:

$$c = z_{(1+\gamma)/2} = z_{1-\alpha/2}. \tag{4.23}$$

Для границ $c = 1$, 2 и 3 получаются значения из таблицы 4.2.

Таблица 4.2. Вероятности центральных гауссовых областей

Центральная область	Вероятность внутри γ	Вероятность каждого хвоста
$\lvert Z\rvert < 1$	0.6827	0.1587
$\lvert Z\rvert < 2$	0.9545	0.0228
$\lvert Z\rvert < 3$	0.9973	0.00135

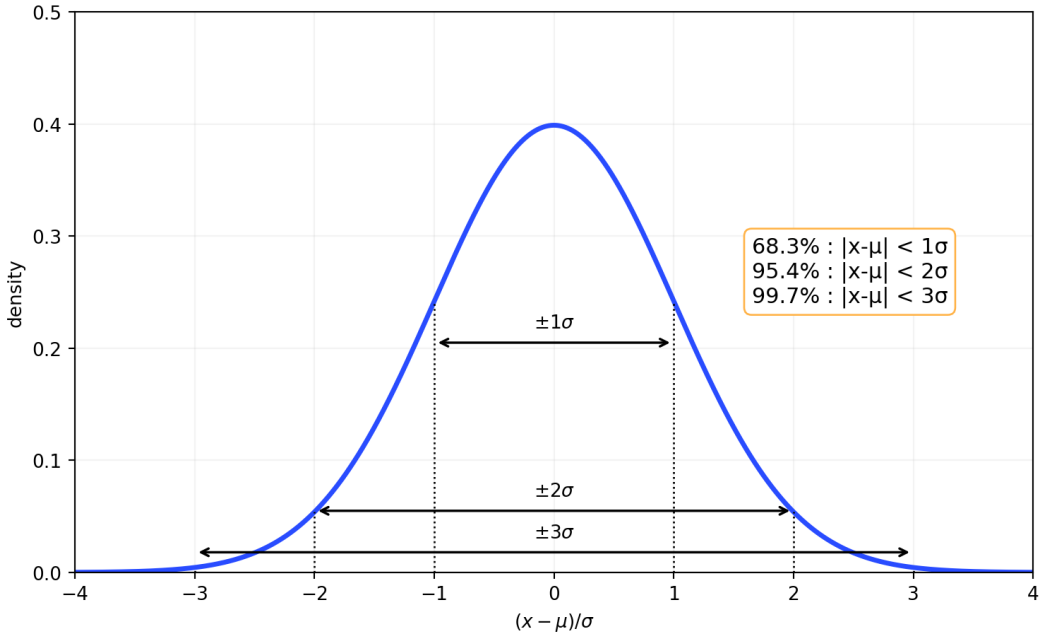


Рис. 4.4. *Плотность стандартного нормального распределения и центральные области в пределах одного, двух и трёх стандартных отклонений*

На рисунке 4.4 границы $\pm k\sigma$ отмечены на одной гауссовой кривой. Запись «интервал 1σ» означает одну σ в каждую сторону от среднего; полная ширина такого интервала равна 2σ.

В физике частиц порог открытия обычно задают односторонней гауссовой значимостью 5σ. Соответствующая хвостовая вероятность равна

$$p_0 = P(Z \geq 5) = 1 - \Phi(5) \approx 2.87 \cdot 10^{-7}. \tag{4.24}$$

Одностороннее и двустороннее соглашения дают разные вероятности при одном числе сигм. Кроме того, локальная хвостовая вероятность меняется после учёта поиска во многих точках параметра, а систематические неопределённости входят в саму статистическую модель. Надпись 5σ не выполняет эти вычисления за исследователя [2].

4.8. Совместимость двух измерений

Пусть два измерения одной физической величины дали

$$x_1 \pm \sigma_1, \qquad x_2 \pm \sigma_2. \tag{4.25}$$

Для начала сравним две пары:

$$1.0 \pm 0.5, \qquad 2.0 \pm 0.5, \tag{4.26}$$

и

$$1.010 \pm 0.001, \qquad 1.020 \pm 0.001. \tag{4.27}$$

В первой паре центральные значения отличаются на единицу, во второй — всего на 0.010. Этого сравнения недостаточно. Масштаб разности задают её собственные флуктуации.

4.8.1. Распределение разности

Обозначим результаты ещё не выполненных измерений через X_1 и X_2 и введём

$$D = X_1 - X_2. \tag{4.28}$$

При гипотезе, что оба эксперимента измеряют одно истинное значение μ,

$$\mathbb{E}[D] = \mu - \mu = 0.$$

Из общей формулы для дисперсии суммы следует

$$\operatorname{Var}(D) = \sigma_1^2 + \sigma_2^2 - 2\operatorname{cov}(X_1, X_2). \tag{4.29}$$

Если измерения независимы, ковариация равна нулю и

$$\sigma_D = \sqrt{\sigma_1^2 + \sigma_2^2}. \tag{4.30}$$

Для независимых гауссовых измерений разность также гауссова. Полученное значение разности переводится в стандартную нормальную шкалу:

$$z_{\text{obs}} = \frac{x_1 - x_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}. \tag{4.31}$$

Двусторонняя хвостовая вероятность для такой или ещё большей по модулю разности равна

$$p = 2\left[1 - \Phi(|z_{\text{obs}}|)\right]. \tag{4.32}$$

Это p-значение относится к конкретной гипотезе: два измерения имеют одно среднее, их стандартные отклонения известны, а ошибки независимы и гауссовы.

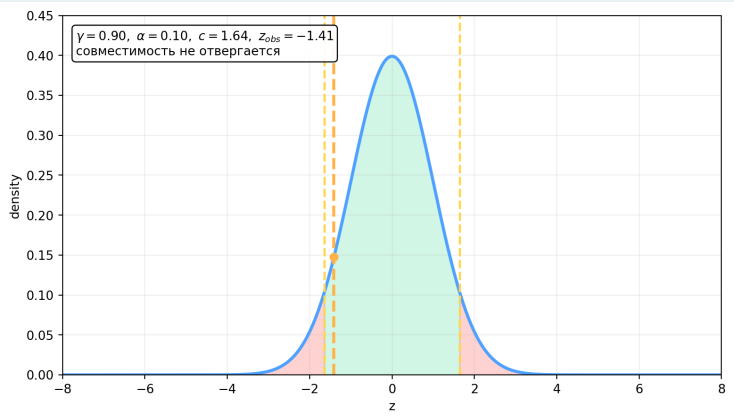
4.8.2. Центральная область проверки

Выберем уровень значимости α. Центральная область неотвержения имеет границу

$$c = z_{1-\alpha/2}. \tag{4.33}$$

Если $|z_{\text{obs}}| \le c$, гипотеза совместимости не отвергается. При $|z_{\text{obs}}| > c$ результат попадает в один из двух хвостов, и гипотеза отвергается на выбранном уровне α. Та же проверка записывается короче: $p \ge \alpha$ или $p < \alpha$.

Фраза «совместимость не отвергается» выбрана намеренно. Она не доказывает, что оба измерения правильны. Два эксперимента могут иметь общий неучтённый сдвиг и прекрасно соглашаться друг с другом.



Интерактив. Совместимость двух измерений
Изменяйте уровень α и полученное значение z_{obs}. На графике показаны центральная область неотвержения, два хвоста и двустороннее p-значение.

4.8.3. Два численных примера

Для пары из уравнения 4.26

$$z_{\text{obs}} = \frac{1.0 - 2.0}{\sqrt{0.5^2 + 0.5^2}} \approx -1.41, \qquad p \approx 0.157. \tag{4.34}$$

При $\alpha = 0.10$ критическая граница равна $z_{0.95} \approx 1.645$. Полученное значение лежит внутри центральной области, и совместимость не отвергается.

Для пары из уравнения 4.27

$$z_{\text{obs}} = \frac{1.010 - 1.020}{\sqrt{0.001^2 + 0.001^2}} \approx -7.07, \qquad p \approx 1.5 \cdot 10^{-12}. \tag{4.35}$$

Разность здесь в сто раз меньше, чем в первом примере, и примерно в пять раз больше в единицах своей неопределённости. Физический смысл имеет второе сравнение.

4.8.4. Общие неопределённости

Независимость двух измерений часто оказывается приближением. Эксперименты могут использовать одну внешнюю константу, общий расчёт потока или одну модель фона. Тогда в уравнении 4.29 остаётся ковариационный член.

Положительная ковариация уменьшает дисперсию разности: общий сдвиг сокращается при вычитании. Отрицательная ковариация её увеличивает. Для совместно гауссовых X_1 и X_2 разность остаётся гауссовой, и формула 4.31 обобщается заменой знаменателя на

$$\sqrt{\sigma_1^2 + \sigma_2^2 - 2\,\mathrm{cov}(X_1, X_2)}. \tag{4.36}$$

Если сами неопределённости оценены по небольшим выборкам или распределения имеют заметные негауссовы хвосты, стандартная нормальная шкала

требует отдельной проверки.

Итоги главы

- Среднее суммы равно сумме средних. В её дисперсию входят дисперсии отдельных вкладов и все попарные ковариации.
- Для независимых одинаково распределённых величин с конечной ненулевой дисперсией центрированная и нормированная сумма сходится к стандартному нормальному распределению. Скорость сходимости зависит от исходного закона, особенно при работе с хвостами.
- Стандартизация переводит отклонение в единицы σ, а квантили задают односторонние и центральные области с выбранной вероятностью.
- Совместимость двух гауссовых измерений определяется распределением их разности. Общие источники неопределённости входят в её дисперсию через ковариацию.

4.9. Задачи

Задача 1. Квантили и отклонения в сигмах

Пусть

$$X \sim \mathrm{N}(12.0, 0.8^2).$$

Найдите:

1. центральный интервал 68%;
2. центральный интервал 95%;
3. односторонний квантиль уровня 0.95;
4. на сколько сигм отклоняется значение $x = 13.4$ от среднего.

Задача 2. Совместимость независимых измерений

Три независимых измерения дали:

$$x_1 = 125.12 \pm 0.30, \quad x_2 = 124.86 \pm 0.18, \quad x_3 = 125.05 \pm 0.24.$$

Для каждой из трёх пар вычислите z_{ij} и двустороннее p-значение. Какие пары совместимы на уровне значимости $\alpha = 0.05$? Сравните порядок пар по абсолютной разности $|x_i - x_j|$ и по $|z_{ij}|$.

Задача 3. Фотоэлектронная статистика

Заряд одного фотоэлектронного импульса имеет среднее $q_0 = 1.0\,\text{пКл}$ и стандартное отклонение $s_q = 0.4\,\text{пКл}$. Импульсы независимы.

1. Найдите средний полный заряд, его стандартное отклонение и относительную ширину для $N = 25$ и $N = 100$ фотоэлектронов.
2. Во сколько раз нужно увеличить N, чтобы относительная ширина уменьшилась вдвое?
3. Какое распределение полного заряда ожидается при большом фиксированном N? Запишите нормированную величину, к которой применяется ЦПТ.

Литература

[1] Feller, W. *An Introduction to Probability Theory and Its Applications*. John Wiley & Sons. 1971.
[2] Cowan, G. *Statistical Data Analysis*. Oxford University Press. 1998.

Глава 5

Когда гаусс не возникает

Содержание

Аннотация

В прошлой главе мы увидели, как сумма независимых случайных вкладов приходит к распределению Гаусса. Теперь разберём случаи, в которых этот вывод применять нельзя или до гауссова предела приходится ждать слишком долго.
Начнём с распределений Коши и Ландау, у которых нет конечной дисперсии. Затем посмотрим, как корреляции меняют закон убывания ошибки, чем смесь распределений отличается от суммы и почему отношение двух гауссовых величин может иметь длинные хвосты.

5.1. Физическая задача: ионизационные потери в трековом детекторе

Заряженная частица проходит через трековый детектор и оставляет сигнал в n последовательных слоях. По измеренным зарядам

$$Q_1, Q_2, \dots, Q_n \tag{5.1}$$

нужно оценить её удельные потери энергии dE/dx. Первая идея состоит в том, чтобы взять среднее:

$$\overline{Q}_n = \frac{1}{n}\sum_{i=1}^{n} Q_i. \tag{5.2}$$

Если Q_i независимы и имеют конечную дисперсию σ_Q^2, то

$$\operatorname{Var}(\overline{Q}_n) = \frac{\sigma_Q^2}{n}. \tag{5.3}$$

Можно ожидать, что с ростом числа слоёв распределение среднего станет узким и гауссовым. Однако распределение заряда в одном тонком слое имеет длинный правый хвост. Иногда частица передаёт большую энергию одному электрону, и один Q_i оказывается намного больше остальных. Обычное среднее полностью сохраняет этот вклад.

В газовых трековых детекторах часто используют **усечённое среднее**. Заряды сначала располагают по возрастанию,

$$Q_{(1)} \le Q_{(2)} \le \dots \le Q_{(n)}, \tag{5.4}$$

а затем усредняют только m меньших значений:

$$\overline{Q}_{\mathrm{tr}} = \frac{1}{m}\sum_{i=1}^{m} Q_{(i)}, \qquad m < n. \tag{5.5}$$

Так редкие большие потери не определяют оценку dE/dx. Выбор доли отбрасываемых измерений зависит от детектора и оптимизируется по его разрешению. Причину длинного хвоста мы получим ниже из спектра передачи энергии электрону [1].

Этот пример сразу ставит правильный вопрос. Число слагаемых само по себе ещё не определяет форму суммы. Нужно знать распределение отдельных вкладов и связи между ними.

5.2. Что именно утверждает центральная предельная теорема

Пусть

$$S_n = X_1 + X_2 + \cdots + X_n. \tag{5.6}$$

Для независимых одинаково распределённых величин с параметрами

$$\mathbb{E}[X_i] = \mu, \qquad 0 < \operatorname{Var}(X_i) = \sigma^2 < \infty \tag{5.7}$$

центральная предельная теорема даёт

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{S_n/n - \mu}{\sigma/\sqrt{n}} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1). \tag{5.8}$$

В этой формулировке содержатся четыре условия, которые понадобится проверять.

- Рассматривается сумма случайных величин.
- Слагаемые независимы и имеют одно распределение.
- Их дисперсия конечна.
- Переход к пределу выполняется после центрирования и нормировки суммы.

Для независимых величин с разными распределениями существуют более общие варианты теоремы. Условия Линдеберга и Ляпунова запрещают одному слагаемому сохранять конечную долю полной дисперсии. Для некоторых зависимых последовательностей также доказаны свои предельные теоремы [2]. Но из одной фразы «слагаемых много» ни одна из них не следует.

Общая формула для дисперсии суммы имеет вид

$$\operatorname{Var}(S_n) = \sum_{i=1}^{n} \operatorname{Var}(X_i) + 2\sum_{i<j} \operatorname{cov}(X_i, X_j). \tag{5.9}$$

При независимых слагаемых ковариации равны нулю. Корреляции оставляют вторую сумму и могут полностью изменить зависимость ширины от n. Если дисперсия X_i не существует, нормировка через $\sqrt{n}$ лишается основания.

5.3. Распределение Коши

Стандартное распределение Коши имеет плотность

$$f_X(x) = \frac{1}{\pi(1 + x^2)}. \tag{5.10}$$

Его кумулятивная функция распределения равна

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi}\arctan x. \tag{5.11}$$

Поэтому величину с распределением Коши можно получить из равномерной $U \sim U(0,1)$ преобразованием

$$X = F_X^{-1}(U) = \tan\left[\pi\left(U - \frac{1}{2}\right)\right]. \tag{5.12}$$

Хвост плотности убывает как $1/x^2$. Из-за этого

$$\mathbb{E}[|X|] = \infty, \qquad \mathbb{E}[X^2] = \infty. \tag{5.13}$$

Математическое ожидание распределения Коши не существует. Симметричное главное значение интеграла равно нулю, но математическим ожиданием оно не является.

5.3.1. Среднее Коши не сужается

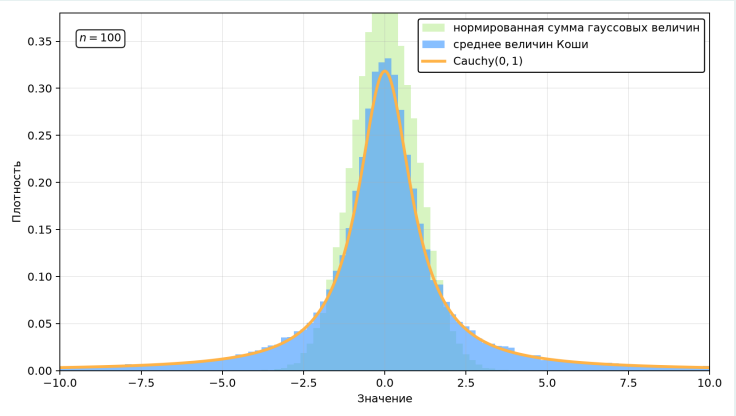
Пусть $X_1, \ldots, X_n$ — независимые стандартные величины Коши. В главе 3 мы нашли характеристическую функцию их суммы:

$$\varphi_{S_n}(t) = e^{-n|t|}, \qquad S_n = \sum_{i=1}^{n} X_i. \tag{5.14}$$

Это характеристическая функция распределения Коши с масштабом n. Деление суммы на n возвращает исходный масштаб:

$$\frac{S_n}{n} \sim \text{Cauchy}(0, 1). \tag{5.15}$$

Распределение среднего при $n = 1$ и $n = 200$ совпадает. Для гауссовых слагаемых ширина среднего уменьшилась бы в $\sqrt{200}$ раз.



Интерактив. Среднее распределения Коши
Меняя число слагаемых, можно сравнить среднее величин Коши с нормированной суммой гауссовых величин.

Распределение Коши встречается и в физике. Нерелятивистская линия Брейта—Вигнера имеет ту же форму и служит простой моделью резонансного пика. Для релятивистского резонанса форма зависит от выбранной кинематической переменной и фазового пространства.

Из поведения среднего не следует, что накопление данных для распределения Коши бесполезно. Параметр положения можно оценивать по медиане или методом максимального правдоподобия. Эти оценки используют форму распределения и не требуют существования $\mathbb{E}[X]$.

5.4. Энергетические потери в тонком слое вещества

Вернёмся к задаче, с которой началась глава. Заряженная частица теряет энергию во множестве актов ионизации и возбуждения. Полную потерю энергии в слое

запишем как

$$\Delta = \omega_1 + \omega_2 + \cdots + \omega_N, \tag{5.16}$$

где ω_a — энергия, переданная в одном столкновении. Формально перед нами сумма. Форма распределения определяется хвостом отдельных ω_a.

5.4.1. Редкие большие передачи энергии

Для быстрой тяжёлой заряженной частицы дифференциальное сечение передачи энергии электрону в существенной для хвоста области ведёт себя как

$$\frac{d\sigma}{d\omega} \propto \frac{1}{\omega^2}, \qquad \omega \le W_{\max}. \tag{5.17}$$

Здесь $W_{\max}$ — максимальная передаваемая энергия, заданная кинематикой столкновения. Вероятность передачи больше порога W пропорциональна

$$P(\omega > W) \propto \int_W^{W_{\max}} \frac{d\omega}{\omega^2} = \frac{1}{W} - \frac{1}{W_{\max}}. \tag{5.18}$$

При $W \ll W_{\max}$ получаем

$$P(\omega > W) \propto \frac{1}{W}. \tag{5.19}$$

Хвост убывает медленно. Один δ-электрон может унести заметную часть энергии, доступной в слое, и дать заряд намного больше наиболее вероятного. Именно эти события образуют правый хвост распределения Q_i в трековом детекторе.

5.4.2. Распределение Ландау

В идеализированном пределе тонкого слоя распределение потери энергии имеет ландау-подобную форму

$$p(\Delta) = \frac{1}{\xi} L\!\left(\frac{\Delta - \Delta_0}{\xi}\right). \tag{5.20}$$

Параметр Δ_0 задаёт положение распределения, а ξ — масштаб флуктуаций. При используемом здесь определении стандартной функции Ландау Δ_0 не совпадает с наиболее вероятным значением. Для частицы с зарядом ze в веществе

$$\xi = \frac{K}{2}\frac{Z}{A}\frac{z^2}{\beta^2}\rho x. \tag{5.21}$$

Здесь Z и A — атомный номер и молярная масса вещества, ρx — массовая толщина слоя, $\beta = v/c$, а $K = 4\pi N_A r_e^2 m_e c^2 \simeq 0.307$ МэВ см^2 моль$^{-1}$.

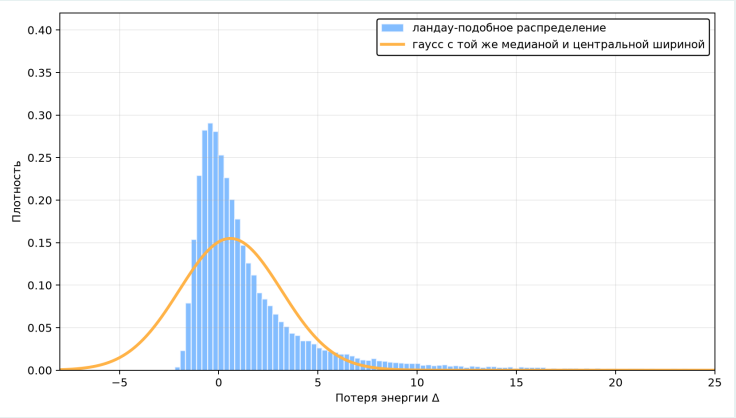
Стандартную функцию Ландау можно определить интегралом

$$L(\lambda) = \frac{1}{\pi}\int_0^{\infty} \exp[-u\ln u - \lambda u]\sin(\pi u)\,du. \tag{5.22}$$

При больших положительных λ

$$L(\lambda) \sim \frac{1}{\lambda^2}. \tag{5.23}$$

Наиболее вероятное значение конечно, а среднее и дисперсия идеального распределения Ландау не определены. Подробное описание флуктуаций ионизационных потерь в тонких кремниевых детекторах дано в [1].



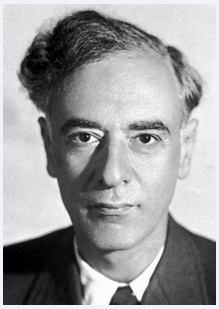
Интерактив. Длинный хвост распределения Ландау
Интерактив сопоставляет ландау-подобное распределение с гауссовой кривой, имеющей ту же медиану и ту же центральную ширину.

В идеализированной модели верхнюю границу передачи энергии удаляют. Тогда хвост $p(\omega) \propto 1/\omega^2$ даёт логарифмически расходящийся первый момент и линейно расходящийся второй:

$$\int^{\infty} \omega \frac{d\omega}{\omega^2} = \int^{\infty} \frac{d\omega}{\omega}, \qquad \int^{\infty} \omega^2 \frac{d\omega}{\omega^2} = \int^{\infty} d\omega. \tag{5.24}$$

Обычная ЦПТ к такой модели неприменима. В реальном столкновении $W_{\max}$ конечно. Распределение Ландау, столь любимое экспериментаторами для описания потерь энергии в тонких детекторах - это приближение. Если хотите точнее, используйте распределение Вавилова. При увеличении толщины слоя последнее переходит к гауссову распределению.

Теперь понятна и конструкция из начала главы. Усечённое среднее подавляет редкие большие Q_i, созданные δ-электронами. Цена этой операции — потеря части измерений и необходимость отдельно определить распределение и смещение оценки $\overline{Q}_{\text{tr}}$.



БИОГРАФИЯ

Лев Давидович Ландау

1908–1968

Лев Ландау - выдающийся советский физик-теоретик. Он разработал квантовую теорию диамагнетизма, теорию фазовых переходов второго рода, теорию сверхтекучести жидкого гелия и теорию ферми-жидкости. Вместе с Виталием Гинзбургом он построил феноменологическую теорию сверхпроводимости. За работы по теории конденсированного состояния, прежде всего жидкого гелия, Ландау получил Нобелевскую премию по физике 1962 года.

Для этой главы особенно важна его теория флуктуаций энергетических потерь быстрой заряженной частицы в тонком слое вещества. Полученное распределение сохраняет след редких больших передач энергии и имеет длинный правый хвост.

Лев Ландау и Евгений Лифшиц создали многотомный «Курс теоретической физики», воспитавший множество поколений учёных в нашей стране и во всём мире.

Источник: Nobel Prize; фото архив Нобелевского фонда, общественное достояние

5.5. Коррелированные слагаемые

Ковариационная сумма в уравнении 5.9 исчезает только для некоррелированных слагаемых. Рассмотрим простой случай:

$$\mathrm{Var}(X_i) = \sigma^2, \qquad \mathrm{corr}(X_i, X_j) = \rho, \quad i \neq j. \tag{5.25}$$

Для среднего

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{5.26}$$

получаем

$$\begin{aligned} \mathrm{Var}(\overline{X}_n) &= \frac{1}{n^2}\left[n\sigma^2 + 2\frac{n(n-1)}{2}\rho\sigma^2\right] \\ &= \frac{\sigma^2}{n}\left[1 + (n-1)\rho\right]. \end{aligned} \tag{5.27}$$

При $\rho = 0$ остаётся σ^2/n. При фиксированном $\rho > 0$

$$\mathrm{Var}(\overline{X}_n) \xrightarrow[n\to\infty]{} \rho\sigma^2. \tag{5.28}$$

Матрица с одинаковой корреляцией всех пар допустима при $-1/(n-1) \le \rho \le 1$. Поэтому фиксированное отрицательное ρ нельзя сохранять при неограниченном росте n. Положительная корреляция оставляет конечную ширину среднего.

5.5.1. Общий сдвиг как систематическая ошибка

Пусть каждое измерение содержит одну общую и одну независимую составляющую:

$$X_i = \mu + C + \varepsilon_i. \tag{5.29}$$

В ансамбле повторных калибровок и измерений зададим

$$C \sim \mathcal{N}(0, \sigma_C^2), \qquad \varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2), \tag{5.30}$$

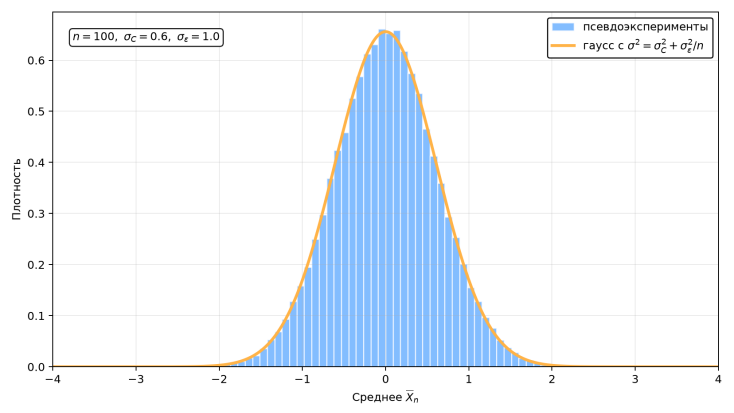
причём C и все ε_i независимы. Тогда

$$\overline{X}_n = \mu + C + \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \tag{5.31}$$

и

$$\mathrm{Var}(\overline{X}_n) = \sigma_C^2 + \frac{\sigma_\varepsilon^2}{n}. \tag{5.32}$$

Статистический член уменьшается с числом измерений. Общий сдвиг повторяется во всех X_i и остаётся целиком. Так ведут себя, например, неопределённость энергетической шкалы, нормировки потока или светимости, коэффициента усиления и общий временной сдвиг.



Интерактив. Статистическая и общая ошибки
Меняйте статистическую и общую составляющие ошибки и следите за шириной распределения среднего при росте выборки.

В этой модели $\overline{X}_n$ имеет гауссово распределение при любом n. Однако его ширина стремится к σ_C и не обращается в нуль. Гауссова форма гистограммы сама по себе не подтверждает закон $1/\sqrt{n}$.

5.6. Смеси распределений

Сумма и смесь описывают разные устройства эксперимента. Для суммы два вклада присутствуют в каждом событии. Например, если

$$N_1 \sim \mathrm{Pois}(1000), \qquad N_2 \sim \mathrm{Pois}(100) \tag{5.33}$$

независимы, то

$$N_1 + N_2 \sim \mathrm{Pois}(1100). \tag{5.34}$$

В смеси каждое событие принадлежит одному из нескольких классов. Введём скрытую переменную $K \in \{1, 2\}$ и зададим

$$N \mid K = 1 \sim \text{Pois}(1000), \qquad N \mid K = 2 \sim \text{Pois}(100). \tag{5.35}$$

Если класс события не зарегистрирован, в данных остаётся только N. Даже узкие распределения внутри классов образуют широкую или многопиковую смесь.

5.6.1. Среднее и дисперсия смеси

Пусть X при фиксированном классе K имеет среднее μ_K и дисперсию σ_K^2. Полное среднее равно

$$\mathbb{E}[X] = \mathbb{E}_K[\mu_K]. \tag{5.36}$$

Формула полной дисперсии разделяет два источника разброса:

$$\text{Var}(X) = \mathbb{E}_K\left[\sigma_K^2\right] + \text{Var}_K(\mu_K). \tag{5.37}$$

Первое слагаемое усредняет дисперсии внутри классов. Второе измеряет разброс между их средними.

Для двух равновероятных пуассоновских классов из уравнения 5.35

$$\mathbb{E}[N] = \frac{1000 + 100}{2} = 550. \tag{5.38}$$

Средняя дисперсия внутри классов равна

$$\mathbb{E}_K[\text{Var}(N \mid K)] = \frac{1000 + 100}{2} = 550, \tag{5.39}$$

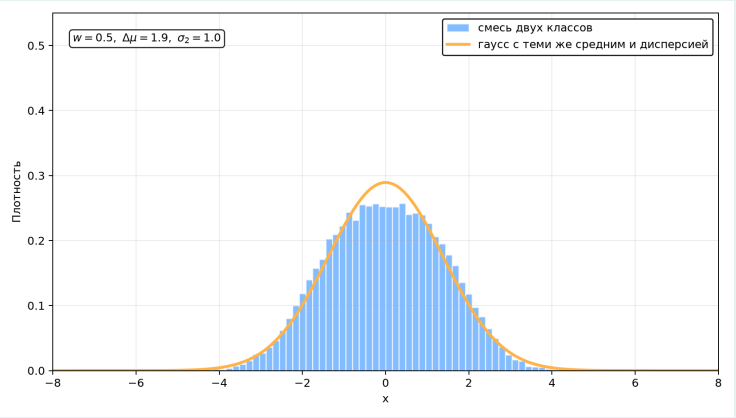
а дисперсия условных средних составляет

$$\text{Var}_K(\mathbb{E}[N \mid K]) = \frac{(1000 - 550)^2 + (100 - 550)^2}{2} = 202\,500. \tag{5.40}$$

Полная дисперсия

$$\text{Var}(N) = 550 + 202\,500 = 203\,050 \tag{5.41}$$

почти целиком определяется различием двух классов. Пуассоновские флуктуации внутри классов дают меньше трёх десятых процента результата.



Интерактив. Смесь двух гауссовых классов
Интерактив показывает, как веса, ширины и расстояние между средними меняют форму смеси.

5.6.2. Пример: жидкий сцинтиллятор

Сигнал жидкосцинтилляционного детектора можно схематично записать как

$$Q = Q(E, K) + Q_{\text{stat}} + Q_{\text{el}}, \tag{5.42}$$

где Q_{stat} описывает фотонную статистику, Q_{el} — шум электроники, а класс K включает тип частицы, топологию события, положение в детекторе, долю черенковского света и другие детали отклика.

Для γ-кванта энергия передаётся нескольким вторичным электронам:

$$\gamma \longrightarrow e_1 + e_2 + \cdots + e_m, \qquad E_\gamma = \sum_{a=1}^{m} E_a. \tag{5.43}$$

Если световой выход электрона $L(E)$ нелинеен, средний заряд имеет вид

$$Q_\gamma \simeq \sum_{a=1}^{m} L(E_a). \tag{5.44}$$

При одном и том же E_γ разные наборы E_a дают разные значения Q_γ. Разброс топологий добавляет к фотонной статистике второй член из формулы полной дисперсии 5.37. Описание сигнала одним пуассоновским или гауссовым распределением этот вклад потеряет.

5.7. Отношение случайных величин

Центральная предельная теорема относится к суммам. Нелинейное преобразование может изменить распределение даже при точно гауссовых исходных величинах.

Пусть X и Y независимы и

$$X, Y \sim \mathcal{N}(0, 1). \tag{5.45}$$

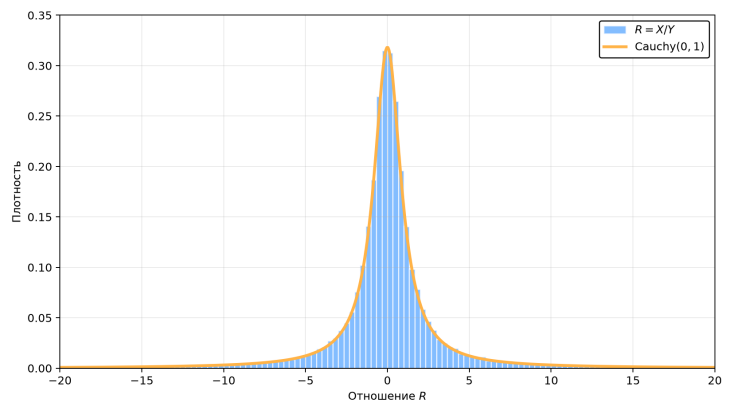
Для отношения $R = X/Y$ совместная плотность после замены $x = ry$ даёт

$$
\begin{aligned}
f_R(r) &= \int_{-\infty}^{+\infty} |y| f_X(ry) f_Y(y)\, dy \\
&= \frac{1}{2\pi} \int_{-\infty}^{+\infty} |y| \exp\left[-\frac{(1+r^2)y^2}{2}\right] dy \\
&= \frac{1}{\pi(1+r^2)}.
\end{aligned}
\tag{5.46}
$$

Получилось стандартное распределение Коши. Отношения появляются при нормировке сигнала на контрольную область, введении поправки на эффективность, сравнении двух скоростей счёта и построении асимметрии

$$
A = \frac{N_1 - N_2}{N_1 + N_2}.
\tag{5.47}
$$

Линейное распространение ошибок для отношения применимо, когда знаменатель с высокой вероятностью далёк от нуля и его относительная неопределённость мала. Если знаменатель может оказаться около нуля, редкие значения отношения уходят далеко в хвост. Симметричная запись $r \pm \sigma_r$ перестаёт описывать такое распределение.



Интерактив. Отношение гауссовых величин
Меняйте среднее и ширину знаменателя и следите за появлением длинных хвостов у отношения.

5.8. Причины отклонения от гауссова приближения

В таблице 5.1 собраны рассмотренные причины.

Таблица 5.1. Причины отклонения от простого гауссова приближения

Механизм	Какое условие требует проверки	Что видно в данных
Бесконечная дисперсия	Существует ли $\mathrm{Var}(X_i)$	Среднее может не сужаться
Тяжёлый хвост	Насколько быстро наступает предельный режим	Редкие большие значения сохраняются при конечном n

Механизм	Какое условие требует проверки	Что видно в данных
Доминирующий вклад	Мал ли каждый вклад по сравнению со всей суммой	Форму хвоста задаёт один механизм
Корреляции	Независимы ли слагаемые	Ширина не обязана убывать как $1/\sqrt{n}$
Смесь классов	Однородна ли выборка	Возникают дополнительная ширина, асимметрия или несколько пиков
Нелинейное преобразование	Является ли величина суммой	Отношение или асимметрия наследует границы и хвосты знаменателя

Проверять гауссово приближение следует на той случайной величине, которая действительно входит в анализ. Для этого строят псевдоэксперименты из полной модели и сравнивают распределения при нескольких размерах выборки. Центральная часть гистограммы может выглядеть гауссовой раньше хвостов, а именно хвосты определяют вероятности редких отклонений.

Итоги главы

- Конечная дисперсия и отсутствие доминирующего слагаемого входят в условия гауссова предела. У среднего величин Коши ширина не уменьшается с ростом выборки.
- Длинный правый хвост ионизационных потерь создают редкие δ-электроны. В тонком слое используется ландау-подобное описание; кинематическая граница передачи энергии приводит к более точному распределению Вавилова.
- Положительные корреляции оставляют конечный вклад в дисперсию среднего. Общая систематическая ошибка может сохранить гауссову форму и одновременно нарушить закон $1/\sqrt{n}$.
- Смесь классов получает дополнительную дисперсию от различия условных средних. Этот вклад описывается формулой полной дисперсии.
- Отношение двух независимых стандартных гауссовых величин имеет распределение Коши. Для нелинейной величины распределение нужно выводить отдельно.

5.9. Задачи

Задача 1. Среднее распределения Коши

Для независимых $X_i \sim \text{Cauchy}(0,1)$ покажите, что дисперсия не существует, а

$$\overline{X}_n = \frac{1}{n}\sum_i X_i$$

имеет то же распределение Коши. Почему это не противоречит ЦПТ?

Задача 2. Тяжёлый хвост с конечной дисперсией

Пусть

$$p(x) \sim \frac{C}{|x|^{1+\alpha}}.$$

1. При каких α существуют среднее и дисперсия?
2. Почему при α немного больше 2 сходимость к гауссу может быть медленной?

Задача 3. Один доминирующий вклад

Пусть

$$S = X_0 + \epsilon_1 + \cdots + \epsilon_n,$$

где $\epsilon_i \sim \mathrm{N}(0, \sigma^2)$, а X_0 негауссов и

$$\mathrm{Var}(X_0) \gg n\sigma^2.$$

Объясните форму распределения S.

Задача 4. Общая систематическая ошибка

Пусть

$$X_i = \mu + C + \epsilon_i,$$

где $C \sim \mathrm{N}(0, \sigma_C^2)$ и $\epsilon_i \sim \mathrm{N}(0, \sigma_\epsilon^2)$. Найдите $\mathbb{E}[\overline{X}_n]$, $\mathrm{Var}(\overline{X}_n)$ и предел дисперсии при $n \to \infty$.

Задача 5. Усреднение коррелированных измерений

Пусть n величин имеют одинаковую дисперсию σ^2 и одинаковый коэффициент корреляции ρ для любой пары.

1. Получите дисперсию их среднего.
2. Укажите допустимый диапазон ρ при фиксированном n.
3. Рассмотрите случаи $\rho = 0$ и $\rho > 0$ при увеличении n.

Задача 6. Смесь двух гауссов

Пусть

$$X|A \sim \mathrm{N}(\mu_A, \sigma_A^2), \qquad X|B \sim \mathrm{N}(\mu_B, \sigma_B^2),$$

а вероятность класса A равна w. Найдите полные среднее и дисперсию. Когда смесь плохо похожа на один гаусс?

Задача 7. Скрытая топология в жидком сцинтилляторе

Моноэнергетический γ-квант с энергией

$$E_\gamma = 1 \text{ МэВ}$$

может передать энергию вторичным электронам двумя способами:

$$\theta = A: \quad T_1 = 1 \text{ МэВ},$$

$$\theta = B: \quad T_1 = T_2 = 0.5 \text{ МэВ}.$$

Из-за нелинейности светового выхода средние сигналы различаются:

$$Q|\theta = A \sim \text{Pois}(1000), \qquad Q|\theta = B \sim \text{Pois}(980).$$

Считайте классы равновероятными.

1. Найдите полное среднее $\mathbb{E}[Q]$.
2. Вычислите среднюю внутриклассовую дисперсию
$$\mathbb{E}_\theta[\text{Var}(Q|\theta)].$$
3. Вычислите топологический вклад
$$\text{Var}_\theta(\mathbb{E}[Q|\theta]).$$
4. Найдите полную дисперсию и стандартное отклонение Q.
5. Сравните результат с пуассоновской моделью
$$Q \sim \text{Pois}(\mathbb{E}[Q]).$$
Какой вклад не описывается этой моделью?

Задача 8. Отношение гауссовых величин

Для независимых $X, Y \sim \text{N}(0,1)$ выведите плотность отношения $R = X/Y$ и покажите, что это распределение Коши.

Литература

[1] Bichsel, H. Straggling in Thin Silicon Detectors. *Reviews of Modern Physics*, **60**, 663–699. 1988. doi:10.1103/RevModPhys.60.663.
[2] Feller, W. *An Introduction to Probability Theory and Its Applications.* John Wiley & Sons. 1971.

РАЗДЕЛ КНИГИ

II. Моделирование и оценивание

Главы

Глава 6

Распространение ошибок

Содержание

Аннотация

Физический результат обычно вычисляется по нескольким величинам, которые измеряет детектор. Вместе с результатом требуется найти его неопределённость. В этой главе мы получим линейный закон распространения ошибок, проверим его на нелинейной функции и запишем в матричном виде с учётом корреляций. Отдельный раздел посвящён разбросу оценок дисперсии и стандартного отклонения.

6.1. Физическая задача: измерение сечения

Эксперимент зарегистрировал $N = 1000$ сигнальных событий. Эффективность отбора равна

$$\varepsilon = 0.80 \pm 0.02, \tag{6.1}$$

а интегральная светимость составляет

$$L = (100 \pm 2)\ \text{фб}^{-1}. \tag{6.2}$$

В упрощённой модели фон отсутствует, число событий имеет распределение Пуассона, а эффективность и светимость измерены независимо от N и друг от друга. Оценка сечения процесса равна

$$\hat{\sigma}_{\text{proc}} = \frac{N}{\varepsilon L} = 12.5\ \text{фб}. \tag{6.3}$$

Для пуассоновского счёта стандартное отклонение числа событий можно оценить как $\sigma_N = \sqrt{N}$. Относительная неопределённость сечения в линейном приближении тогда равна

$$\begin{aligned} \left(\frac{\sigma_{\hat{\sigma}_{\text{proc}}}}{\hat{\sigma}_{\text{proc}}}\right)^2 &\simeq \left(\frac{\sigma_N}{N}\right)^2 + \left(\frac{\sigma_\varepsilon}{\varepsilon}\right)^2 + \left(\frac{\sigma_L}{L}\right)^2 \\ &= \left(\frac{\sqrt{1000}}{1000}\right)^2 + \left(\frac{0.02}{0.80}\right)^2 + \left(\frac{2}{100}\right)^2 = 0.045^2. \end{aligned} \tag{6.4}$$

Результат можно записать как

$$\hat{\sigma}_{\text{proc}} = (12.5 \pm 0.6)\ \text{фб}. \tag{6.5}$$

В одной строке под корнем встретились статистика событий, эффективность детектора и светимость ускорителя. Формулу мы пока использовали как готовую. Ниже выведем её и увидим, где в ней должны стоять ковариации. В реальном анализе N получают после вычитания фона, а эффективность и нормировки могут зависеть от общих калибровок. Тогда три квадрата уже не исчерпывают задачу.

6.2. Что называется ошибкой

В названии главы используется привычное физикам выражение «распространение ошибок». В строгой метрологической терминологии ошибка — это разность между измеренным и опорным значениями, а неопределённость описывает разброс возможных результатов. В формулах этой главы речь идёт о стандартной неопределённости, которую мы обозначаем через σ_X.

Стандартное отклонение описывает масштаб случайного разброса. Смещение метода, неопределённость калибровки и физические границы параметра требуют отдельного учёта. Общие правила представления неопределённостей измерения собраны в руководстве JCGM [1].

6.3. Одна измеряемая величина

Пусть детектор измеряет случайную величину X, а результат вычисляется по формуле

$$Y = f(X). \tag{6.6}$$

Обозначим

$$\mathbb{E}[X] = \mu_X, \qquad \operatorname{Var}(X) = \sigma_X^2. \tag{6.7}$$

6.3.1. Линейное распространение ошибки

На масштабе σ_X функцию f можно заменить её касательной в точке μ_X, если кривизна функции на этом интервале мала.

Вывод

6.3.1.1. Формула для одной переменной

Первый порядок разложения Тейлора даёт

$$f(X) \simeq f(\mu_X) + f'(\mu_X)(X - \mu_X). \tag{6.8}$$

В этом приближении

$$\mathbb{E}[Y] \simeq f(\mu_X), \qquad Y - \mathbb{E}[Y] \simeq f'(\mu_X)(X - \mu_X). \tag{6.9}$$

Квадрат производной можно вынести из математического ожидания:

$$\begin{aligned} \operatorname{Var}(Y) &\simeq \mathbb{E}\big[f'(\mu_X)^2 (X - \mu_X)^2\big] \\ &= f'(\mu_X)^2 \sigma_X^2. \end{aligned} \tag{6.10}$$

Поэтому

$$\boxed{\sigma_Y \simeq |f'(\mu_X)| \sigma_X.} \tag{6.11}$$

Производная в уравнении 6.11 показывает, во сколько раз небольшое отклонение входной величины растягивается или сжимается при переходе к Y. Для линейной функции формула точна. Для нелинейной её точность определяется размером области, занятой распределением X.

6.3.2. Поперечный импульс в магнитном поле

Радиус R траектории заряженной частицы в однородном магнитном поле связан с поперечным импульсом соотношением

$$p_T = |q| B R. \tag{6.12}$$

В единицах, обычных для физики частиц,

$$p_T[\text{ГэВ}/c] \simeq 0.2998 \, |q/e| \, B[\text{Тл}] \, R[\text{м}]. \tag{6.13}$$

Если q и B считаются известными, неопределённость создаёт измерение радиуса. Поскольку зависимость линейна,

$$\sigma_{p_T} = |q|B\sigma_R, \qquad \frac{\sigma_{p_T}}{p_T} = \frac{\sigma_R}{R}. \tag{6.14}$$

Ошибка радиуса переходит в такую же относительную ошибку импульса. Если неопределённость магнитного поля существенна, B становится второй входной величиной. Такой случай разберём в разделе о ковариациях.

6.3.3. Степенная зависимость

Для функции

$$Y = X^a \tag{6.15}$$

производная равна

$$f'(x) = ax^{a-1}. \tag{6.16}$$

В точке $x = \mu_X$ уравнение 6.11 принимает вид

$$\frac{\sigma_Y}{|f(\mu_X)|} \simeq |a|\frac{\sigma_X}{|\mu_X|}. \tag{6.17}$$

Несколько часто встречающихся функций собраны в таблице 6.1. Здесь $x_0 = \mu_X$ и $y_0 = f(x_0)$.

Таблица 6.1. Линейное распространение ошибки для часто встречающихся функций

Функция $f(x)$	Линейная неопределённость	Условие на точку разложения
$ax + b$	$\sigma_Y = \lvert a\rvert\sigma_X$	формула точна
x^a	$\sigma_Y/\lvert y_0\rvert \simeq \lvert a\rvert\sigma_X/\lvert x_0\rvert$	функция определена около x_0
$1/x$	$\sigma_Y/\lvert y_0\rvert \simeq \sigma_X/\lvert x_0\rvert$	x_0 далёк от нуля
$\ln x$	$\sigma_Y \simeq \sigma_X/x_0$	$x_0 > 0$, вероятность $X \le 0$ мала
e^x	$\sigma_Y/y_0 \simeq \sigma_X$	X безразмерна

Таблица даёт локальные формулы. У отношения, логарифма и других функций с границей одной малости σ_X недостаточно. Распределение X должно с большой вероятностью оставаться в области, где функция гладкая.

6.4. Где линейное приближение перестаёт работать

Касательная описывает функцию в малой окрестности точки разложения. Размер этой окрестности задаёт σ_X. Линейная формула требует проверки в четырёх типичных ситуациях:

- производная заметно меняется на интервале порядка σ_X;
- $f'(\mu_X)$ близка к нулю и первым ненулевым оказывается квадратичный член;
- рядом находится полюс, порог или физическая граница;

- преобразованное распределение заметно асимметрично.

Возьмём $Y = X^2$ и $\mu_X = 0$. Касательная к параболе в нуле горизонтальна, и линейная формула даёт $\sigma_Y \simeq 0$. Сама величина X^2 при этом флуктуирует. Нулевой ответ сообщает о непригодности первого порядка разложения.

6.4.1. Нелинейное преобразование $Y = X + aX^2$

Рассмотрим модель из лекции:

$$X \sim \mathcal{N}(1, \sigma_X^2), \qquad Y = X + aX^2. \tag{6.18}$$

Касательная в точке $\mu_X = 1$ даёт

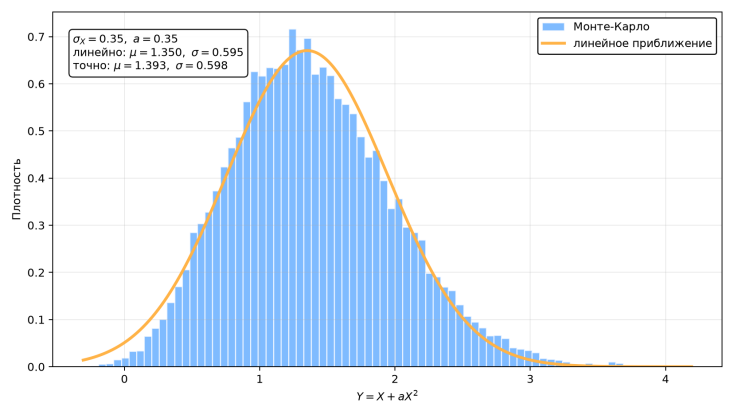
$$\mathbb{E}[Y] \simeq 1 + a, \qquad \sigma_Y \simeq |1 + 2a|\sigma_X. \tag{6.19}$$

Для гауссовой величины моменты можно вычислить без приближения:

$$\mathbb{E}[Y] = 1 + a + a\sigma_X^2, \tag{6.20}$$

$$\mathrm{Var}(Y) = (1 + 2a)^2\sigma_X^2 + 2a^2\sigma_X^4. \tag{6.21}$$

Член $a\sigma_X^2$ сдвигает среднее, а $2a^2\sigma_X^4$ добавляет дисперсию. Оба исчезают в линейном приближении.



Интерактив. Нелинейное преобразование $Y = X + aX^2$
Меняйте ширину исходного гаусса и коэффициент a. Гистограмма показывает распределение Y, а гладкая кривая — линейный прогноз.

На рисунке 6.1 показан режим, в котором кривизна уже существенна на масштабе входного распределения.

Здесь расхождение видно по положению максимума, ширине и длинному правому хвосту. Запись $y \pm \sigma_Y$ сохраняет два числа, но уже не описывает форму распределения. В такой ситуации используют преобразование полного распределения или численное моделирование.

6.5. Разброс оценки стандартного отклонения

До сих пор μ и σ были известными параметрами распределения. В эксперименте их оценивают по выборке

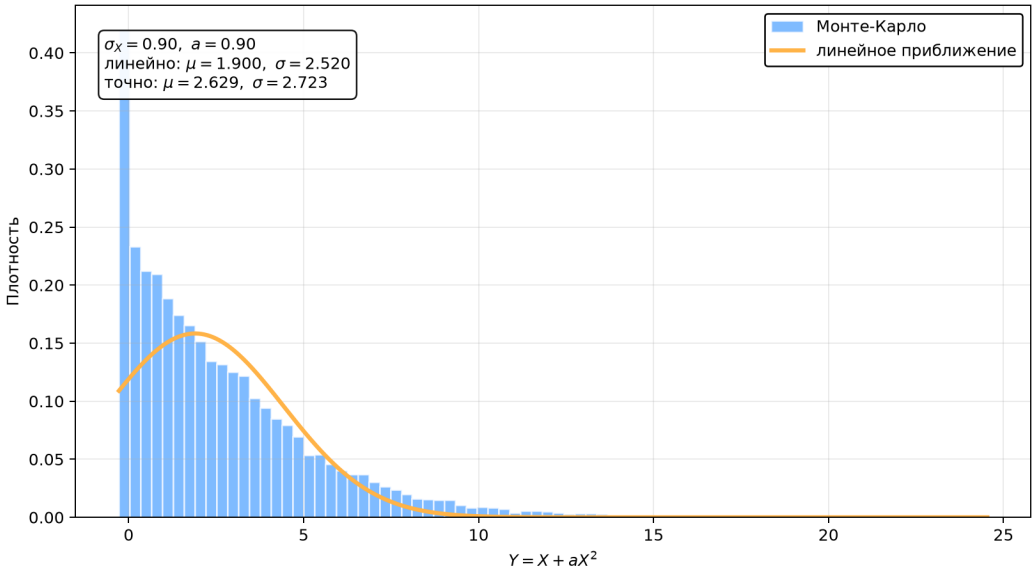


Рис. 6.1. *Распределение $Y = X + aX^2$ при $\sigma_X = 0.9$ и $a = 0.9$. Синяя гистограмма получена методом Монте-Карло, оранжевая кривая показывает гауссово распределение с параметрами линейного приближения.*

$$X_1, \dots, X_N \sim \mathcal{N}(\mu, \sigma^2), \tag{6.22}$$

где измерения независимы. Определим оценки среднего, дисперсии и стандартного отклонения:

$$\hat{\mu} = \overline{X} = \frac{1}{N}\sum_{k=1}^{N} X_k, \tag{6.23}$$

$$\widehat{\sigma^2} = \frac{1}{N-1}\sum_{k=1}^{N}(X_k - \overline{X})^2, \qquad \hat{\sigma} = \sqrt{\widehat{\sigma^2}}. \tag{6.24}$$

Если повторить весь эксперимент, выборка изменится вместе со всеми тремя оценками. Поэтому у $\hat{\sigma}$ есть собственное распределение и собственная дисперсия.

Для выборочного среднего

$$\mathrm{Var}(\hat{\mu}) = \frac{\sigma^2}{N}. \tag{6.25}$$

Параметр σ^2 неизвестен, и в оценке дисперсии среднего его заменяют выборочной дисперсией:

$$\widehat{\mathrm{Var}}(\hat{\mu}) = \frac{\widehat{\sigma^2}}{N}, \qquad \hat{\sigma}_{\hat{\mu}} = \frac{\hat{\sigma}}{\sqrt{N}}. \tag{6.26}$$

Возникает следующий вопрос: насколько устойчиво значение $\hat{\sigma}$, найденное по конечной выборке?

6.5.1. Почему в выборочной дисперсии стоит $N - 1$

Введём отклонения от истинного среднего

$$Y_k = X_k - \mu. \tag{6.27}$$

Для независимой выборки с конечной дисперсией

$$\mathbb{E}[Y_k] = 0, \qquad \mathbb{E}[Y_k^2] = \sigma^2. \tag{6.28}$$

Гауссовость для доказательства несмещённости выборочной дисперсии не нужна.

Вывод

6.5.1.1. Несмещённость выборочной дисперсии

Обозначим две суммы

$$A = \sum_{k=1}^{N} Y_k^2, \qquad B = \sum_{k=1}^{N} Y_k. \tag{6.29}$$

Разность выборочного и истинного среднего равна

$$\overline{X} - \mu = \frac{B}{N}. \tag{6.30}$$

Сумма квадратов отклонений от выборочного среднего принимает вид

$$\begin{aligned} T &= \sum_{k=1}^{N} (X_k - \overline{X})^2 \\ &= \sum_{k=1}^{N} \left(Y_k - \frac{B}{N} \right)^2 \\ &= A - \frac{B^2}{N}. \end{aligned} \tag{6.31}$$

Независимость и центрирование дают

$$\mathbb{E}[A] = N\sigma^2, \qquad \mathbb{E}[B^2] = N\sigma^2. \tag{6.32}$$

Отсюда

$$\mathbb{E}[T] = N\sigma^2 - \frac{1}{N} N\sigma^2 = (N-1)\sigma^2. \tag{6.33}$$

После деления на $N - 1$

$$\boxed{\mathbb{E}[\widehat{\sigma^2}] = \sigma^2.} \tag{6.34}$$

Выборочные отклонения связаны условием

$$\sum_{k=1}^{N} (X_k - \overline{X}) = 0. \tag{6.35}$$

Одно из N отклонений определяется через остальные. Поэтому сумма их квадратов содержит $N - 1$ степеней свободы, и такой же делитель появляется в несмещённой оценке дисперсии.

6.5.2. Дисперсия оценки дисперсии

Теперь понадобится гауссовость. Для центрированной гауссовой величины

$$\mathbb{E}[Y_k^4] = 3\sigma^4. \tag{6.36}$$

Поскольку $\widehat{\sigma^2} = T/(N-1)$, достаточно найти дисперсию T.

Вывод

6.5.2.1. Вычисление $\mathrm{Var}(\widehat{\sigma^2})$

Из уравнения 6.31

$$T^2 = A^2 - \frac{2}{N}AB^2 + \frac{1}{N^2}B^4. \tag{6.37}$$

Первый момент вычисляется раскрытием квадрата суммы:

$$\begin{aligned}\mathbb{E}[A^2] &= \sum_{k=1}^{N}\mathbb{E}[Y_k^4] + 2\sum_{i<j}\mathbb{E}[Y_i^2 Y_j^2] \\ &= 3N\sigma^4 + N(N-1)\sigma^4 \\ &= N(N+2)\sigma^4.\end{aligned} \tag{6.38}$$

Для смешанного момента выделим одно слагаемое, $B = Y_i + C_i$, где $C_i = \sum_{j\neq i} Y_j$. Величины Y_i и C_i независимы, $\mathbb{E}[C_i] = 0$ и $\mathbb{E}[C_i^2] = (N-1)\sigma^2$. Поэтому

$$\begin{aligned}\mathbb{E}[Y_i^2 B^2] &= \mathbb{E}[Y_i^2 (Y_i + C_i)^2] \\ &= 3\sigma^4 + (N-1)\sigma^4 \\ &= (N+2)\sigma^4,\end{aligned} \tag{6.39}$$

и после суммирования по i

$$\mathbb{E}[AB^2] = N(N+2)\sigma^4. \tag{6.40}$$

Сумма B также гауссова, причём $\mathrm{Var}(B) = N\sigma^2$. Её четвёртый момент равен

$$\mathbb{E}[B^4] = 3N^2\sigma^4. \tag{6.41}$$

Подстановка трёх средних в уравнение 6.37 даёт

$$\mathbb{E}[T^2] = (N^2-1)\sigma^4. \tag{6.42}$$

Вместе с уравнением 6.33 это приводит к

$$\mathrm{Var}(T) = 2(N-1)\sigma^4. \tag{6.43}$$

Искомая дисперсия оценки равна

$$\boxed{\mathrm{Var}(\widehat{\sigma^2}) = \frac{2\sigma^4}{N-1}.} \tag{6.44}$$

Уравнение 6.44 является точным для независимой гауссовой выборки. Для другого исходного распределения в ответ входит его четвёртый центральный момент.

6.5.3. Разброс оценки стандартного отклонения

Оценка $\hat{\sigma}$ связана с $\widehat{\sigma^2}$ функцией квадратного корня. Применим к ней формулу 6.11 в точке σ^2:

$$f(y) = \sqrt{y}, \qquad f'(\sigma^2) = \frac{1}{2\sigma}. \tag{6.45}$$

Получаем

$$\mathrm{Var}(\hat{\sigma}) \simeq \frac{1}{4\sigma^2}\,\mathrm{Var}(\widehat{\sigma^2}) = \frac{\sigma^2}{2(N-1)}. \tag{6.46}$$

Относительный разброс оценки стандартного отклонения равен

$$\boxed{\frac{\sqrt{\mathrm{Var}(\hat{\sigma})}}{\sigma} \simeq \frac{1}{\sqrt{2(N-1)}}.} \tag{6.47}$$

Здесь снова использовано линейное приближение, теперь уже для квадратного корня. При конечном N оценка $\hat{\sigma}$ немного смещена вниз; с ростом выборки смещение исчезает, и формула становится точнее.

Для относительного разброса около 10%

$$\frac{1}{\sqrt{2(N-1)}} \simeq 0.1, \qquad N \simeq 51. \tag{6.48}$$

Пятьдесят измерений нужны лишь для того, чтобы сама оценка ширины флуктуировала примерно на десять процентов. Ширину распределения обычно измерить труднее, чем его среднее.

6.5.4. Десять гауссовских измерений

k	x_k	$x_k - \bar{x}$	$(x_k - \bar{x})^2$
1	1.250	0.377	0.142
2	1.008	0.134	0.018
3	1.850	0.976	0.953
4	1.533	0.659	0.435
5	-0.080	-0.953	0.909
6	2.278	1.404	1.972
7	3.534	2.661	7.080
8	-0.146	-1.020	1.040
9	-0.648	-1.521	2.314
10	-1.844	-2.717	7.383

Истинные параметры
$\mu = 0.800,\ \sigma = 1.500,\ N = 10$

Оценки по выборке
$\bar{x} = 0.874$
$\hat{\sigma}^2 = 2.472$
$\hat{\sigma} = 1.572$

Неопределённость среднего
$\hat{\sigma}/\sqrt{N} = 0.497$
$\sqrt{\mathrm{Var}(\bar{x})} = 0.247$

Неопределённость $\hat{\sigma}$
$\hat{\sigma}/\sqrt{2(N-1)} = 0.371$
$\sqrt{\mathrm{Var}(\hat{\sigma})} = 0.137$

Интерактив. Десять гауссовских измерений
Меняйте параметры распределения и начальное состояние генератора. Таблица показывает выборку, оценки $\hat{\mu}$, $\widehat{\sigma^2}$, $\hat{\sigma}$ и их расчётные неопределённости.

Десять чисел выглядят вполне убедительно, пока рядом не написаны истинные параметры. Новая выборка даёт другие $\hat{\mu}$ и $\hat{\sigma}$. Формулы выше описывают разброс этих результатов в серии повторений.

6.6. Несколько измеряемых величин

В задаче о сечении использовались три входные величины. Теперь выведем формулу, которая учитывает их совместные флуктуации.

6.6.1. Ковариация и ошибка суммы

Ковариация X и Y равна

$$\mathrm{cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]. \tag{6.49}$$

Для суммы $S = X + Y$

$$\mathrm{Var}(S) = \sigma_X^2 + \sigma_Y^2 + 2\,\mathrm{cov}(X, Y). \tag{6.50}$$

Коэффициент корреляции убирает размерность ковариации:

$$\rho_{XY} = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y}, \qquad -1 \le \rho_{XY} \le 1. \tag{6.51}$$

Для суммы и разности

$$\sigma_{X\pm Y}^2 = \sigma_X^2 + \sigma_Y^2 \pm 2\rho_{XY}\sigma_X\sigma_Y. \tag{6.52}$$

Положительная корреляция увеличивает ошибку суммы и уменьшает ошибку разности. При отрицательной корреляции знаки меняются. Складывать ошибки в квадратуре можно при нулевой ковариации.

Независимость влечёт $\rho_{XY} = 0$. Обратное утверждение для произвольного распределения неверно: при нулевой корреляции нелинейная зависимость может сохраняться.

6.6.2. Оценка корреляции по выборке

Пусть получены N независимых пар (x_k, y_k) из одного совместного распределения. Выборочная ковариация равна

$$\widehat{\mathrm{cov}}(X, Y) = \frac{1}{N-1}\sum_{k=1}^{N}(x_k - \overline{x})(y_k - \overline{y}), \tag{6.53}$$

а выборочный коэффициент корреляции —

$$\hat{\rho} = \frac{\widehat{\mathrm{cov}}(X, Y)}{s_X s_Y}. \tag{6.54}$$

При небольшом N значение $\hat{\rho}$ заметно меняется от выборки к выборке. В детекторных задачах ковариации получают также из моделирования, калибровочных данных и вариаций систематических параметров.

6.6.3. Общая формула распространения

Пусть

$$F = f(X_1, \dots, X_n), \tag{6.55}$$

а средние входных величин образуют вектор $\mu = (\mu_1, \dots, \mu_n)^{\mathsf{T}}$. Первый порядок разложения имеет вид

$$F - \mathbb{E}[F] \simeq \sum_{i=1}^{n} \left.\frac{\partial f}{\partial x_i}\right|_{\mu} (X_i - \mu_i). \tag{6.56}$$

Вывод

6.6.3.1. Дисперсия функции нескольких переменных

Обозначим производные в точке μ через

$$g_i = \left.\frac{\partial f}{\partial x_i}\right|_{\mu}. \tag{6.57}$$

После возведения линейного разложения в квадрат

$$\begin{aligned}\operatorname{Var}(F) &\simeq \mathbb{E}\left[\left(\sum_i g_i(X_i-\mu_i)\right)\left(\sum_j g_j(X_j-\mu_j)\right)\right] \\ &= \sum_{i,j} g_i \operatorname{cov}(X_i,X_j) g_j.\end{aligned} \tag{6.58}$$

Итак,

$$\boxed{\operatorname{Var}(F) \simeq \sum_{i,j} \frac{\partial f}{\partial x_i} \operatorname{cov}(X_i,X_j) \frac{\partial f}{\partial x_j}.} \tag{6.59}$$

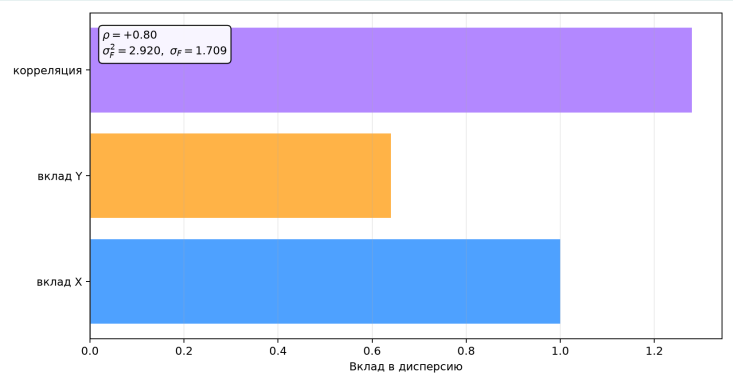
Все производные в уравнении 6.59 вычисляются в одной точке μ. На практике неизвестные средние заменяют полученными центральными значениями.

Для двух переменных формула разворачивается в три слагаемых:

$$\begin{aligned}\sigma_F^2 \simeq &\left(\frac{\partial f}{\partial x}\right)^2 \sigma_X^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_Y^2 \\ &+ 2\rho_{XY} \frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\sigma_X\sigma_Y.\end{aligned} \tag{6.60}$$

Знак корреляционного вклада определяется одновременно знаком ρ_{XY} и знаками двух производных.

6.6.4. Корреляционный вклад в дисперсию



Интерактив. Корреляционный вклад в дисперсию Меняйте производные, стандартные отклонения и ρ. Три полосы показывают два диагональных и один корреляционный вклад в σ_F^2.

На рисунке 6.2 производные имеют одинаковый знак, а корреляция отрицательна. Корреляционный член вычитается из двух положительных вкладов.

6.6.5. Отношение двух величин

Для

$$R = \frac{X}{Y} \tag{6.61}$$

две производные равны

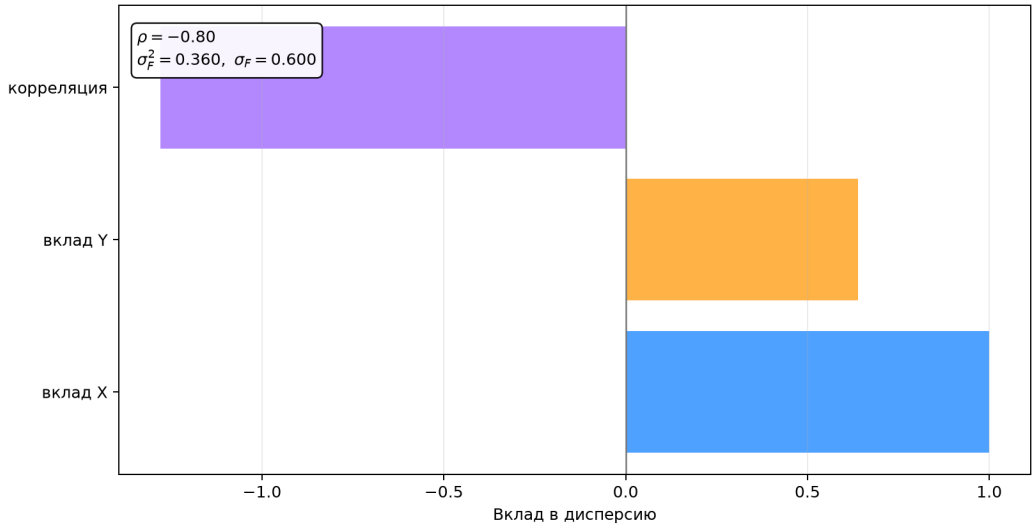


Рис. 6.2. *Вклады в дисперсию функции двух переменных при* $\partial f/\partial x = \partial f/\partial y = 1$, $\sigma_X = 1$, $\sigma_Y = 0.8$ *и* $\rho = -0.8$. *Отрицательный корреляционный член уменьшает полную дисперсию до* 0.36.

$$\frac{\partial R}{\partial X} = \frac{1}{Y}, \qquad \frac{\partial R}{\partial Y} = -\frac{X}{Y^2}. \tag{6.62}$$

Линейная относительная неопределённость составляет

$$\boxed{\left(\frac{\sigma_R}{R}\right)^2 \simeq \left(\frac{\sigma_X}{X}\right)^2 + \left(\frac{\sigma_Y}{Y}\right)^2 - 2\rho_{XY}\frac{\sigma_X}{X}\frac{\sigma_Y}{Y}.} \tag{6.63}$$

Производные по числителю и знаменателю имеют разные знаки. Поэтому положительная корреляция уменьшает ошибку отношения. Формула требует, чтобы знаменатель с большой вероятностью оставался вдали от нуля. В главе 5 мы видели крайний случай: отношение двух независимых стандартных гауссовых величин имеет распределение Коши.

6.6.6. Возвращение к сечению

Для функции

$$\hat{\sigma}_{\text{proc}}(N, \varepsilon, L) = \frac{N}{\varepsilon L} \tag{6.64}$$

производные удобно записать через само сечение:

$$\frac{\partial \hat{\sigma}_{\text{proc}}}{\partial N} = \frac{\hat{\sigma}_{\text{proc}}}{N}, \qquad \frac{\partial \hat{\sigma}_{\text{proc}}}{\partial \varepsilon} = -\frac{\hat{\sigma}_{\text{proc}}}{\varepsilon}, \qquad \frac{\partial \hat{\sigma}_{\text{proc}}}{\partial L} = -\frac{\hat{\sigma}_{\text{proc}}}{L}. \tag{6.65}$$

Общая относительная дисперсия равна

$$\begin{aligned}\left(\frac{\sigma_{\hat{\sigma}_{\text{proc}}}}{\hat{\sigma}_{\text{proc}}}\right)^2 \simeq{}& \frac{\sigma_N^2}{N^2} + \frac{\sigma_\varepsilon^2}{\varepsilon^2} + \frac{\sigma_L^2}{L^2} \\ &- 2\frac{\operatorname{cov}(N,\varepsilon)}{N\varepsilon} - 2\frac{\operatorname{cov}(N,L)}{NL} + 2\frac{\operatorname{cov}(\varepsilon,L)}{\varepsilon L}.\end{aligned} \tag{6.66}$$

При нулевых ковариациях остаётся формула, использованная в начале главы. Знаки трёх ковариационных членов следуют из знаков производных, поэтому механическое сложение всех вкладов в квадратуре может дать неверный ответ.

6.7. Ковариационная матрица

Для случайного вектора

$$\mathbf{X} = (X_1, \dots, X_n)^{\mathsf{T}} \tag{6.67}$$

определим матрицу

$$(\mathbf{V}_X)_{ij} = \operatorname{cov}(X_i, X_j). \tag{6.68}$$

В двумерном случае

$$\mathbf{V}_X = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}. \tag{6.69}$$

Ковариационная матрица симметрична. На диагонали стоят дисперсии, а для любого вектора $\mathbf{a}$

$$\mathbf{a}^{\mathsf{T}}\mathbf{V}_X\mathbf{a} = \operatorname{Var}(\mathbf{a}^{\mathsf{T}}\mathbf{X}) \geq 0. \tag{6.70}$$

Последнее равенство объясняет требование положительной полуопределённости: никакая линейная комбинация измерений не может иметь отрицательную дисперсию. Матрица с нарушенным условием не описывает совместное распределение случайных величин.

6.7.1. Матричная запись распространения ошибки

Соберём производные скалярной функции в градиент

$$\mathbf{g} = \nabla f|_{\mu} = \begin{pmatrix} \partial f/\partial x_1 \\ \vdots \\ \partial f/\partial x_n \end{pmatrix}_{\mu}. \tag{6.71}$$

Уравнение 6.59 записывается одной строкой:

$$\boxed{\sigma_F^2 \simeq \mathbf{g}^{\mathsf{T}}\mathbf{V}_X\mathbf{g}.} \tag{6.72}$$

Пусть выходных величин несколько,

$$\mathbf{Y} = \mathbf{f}(\mathbf{X}), \tag{6.73}$$

а матрица Якоби имеет элементы

$$J_{ai} = \left.\frac{\partial f_a}{\partial x_i}\right|_{\mu}. \tag{6.74}$$

Тогда

$$\boxed{\mathbf{V}_Y \simeq \mathbf{J}\mathbf{V}_X\mathbf{J}^{\mathsf{T}}.} \tag{6.75}$$

Для линейного преобразования матрица $\mathbf{J}$ постоянна и уравнение 6.75 является точным. Для нелинейной функции это первый порядок разложения около μ.

6.7.2. Сумма и разность

Возьмём

$$\mathbf{X} = \begin{pmatrix} X \\ Y \end{pmatrix}, \qquad \mathbf{U} = \begin{pmatrix} S \\ D \end{pmatrix} = \begin{pmatrix} X + Y \\ X - Y \end{pmatrix}. \tag{6.76}$$

Матрица Якоби равна

$$\mathbf{J} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{6.77}$$

Для входной матрицы

$$\mathbf{V}_X = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix} \tag{6.78}$$

получаем

$$\mathbf{V}_U = \mathbf{J}\mathbf{V}_X\mathbf{J}^{\mathsf{T}} = \begin{pmatrix} \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY} & \sigma_X^2 - \sigma_Y^2 \\ \sigma_X^2 - \sigma_Y^2 & \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY} \end{pmatrix}. \tag{6.79}$$

Диагональные элементы воспроизводят ошибки суммы и разности. Внедиагональный элемент даёт

$$\operatorname{cov}(S, D) = \sigma_X^2 - \sigma_Y^2. \tag{6.80}$$

При $\sigma_X = \sigma_Y$ сумма и разность некоррелированы даже при $\sigma_{XY} \neq 0$. В следующей главе мы увидим геометрический смысл этого результата на двумерном гауссовом распределении.

Итоги главы

- При малой входной неопределённости функция заменяется касательной, и для одной переменной $\sigma_Y \simeq |f'(\mu_X)|\sigma_X$.
- Выборочная дисперсия с делителем $N - 1$ является несмещённой. Для гауссовой выборки её дисперсия равна $2\sigma^4/(N-1)$.
- Корреляции входят в ошибку результата со знаками, заданными производными. Квадратурная сумма применима при нулевых ковариациях.
- Ковариационная матрица преобразуется по закону $\mathbf{V}_Y \simeq \mathbf{J}\mathbf{V}_X\mathbf{J}^{\mathsf{T}}$; для линейной функции это равенство точное.

6.8. Задачи

Задача 1. Радиус и площадь круга

Измеренный радиус круга равен

$$R = (10.0 \pm 0.3)\ \text{см}.$$

1. Найдите площадь $S = \pi R^2$ и её ошибку в линейном приближении.
2. Сравните относительные ошибки S и R.
3. Оцените поправку к математическому ожиданию площади, возникающую из-за квадратичности функции.

Задача 2. Сумма и разность коррелированных величин

Пусть

$$\sigma_X = 2, \qquad \sigma_Y = 3, \qquad \rho_{XY} = 0.8.$$

Вычислите ошибки:

$$S = X + Y, \qquad D = X - Y.$$

Повторите вычисление для $\rho_{XY} = -0.8$ и объясните изменение результата.

Задача 3. Ошибка отношения

Эффективность определяется отношением

$$\varepsilon = \frac{N_{\text{selected}}}{N_{\text{total}}}.$$

1. Выведите линейную формулу для σ_ε, предполагая известные σ_{selected}, σ_{total} и корреляцию ρ.
2. Объясните, почему предположение о независимости числителя и знаменателя обычно неверно.
3. Сравните ответы для $\rho = 0$ и $\rho > 0$.

Задача 4. Преобразование ковариационной матрицы

Пусть

$$\mathbf{X} = \begin{pmatrix} X \\ Y \end{pmatrix}, \qquad V_X = \begin{pmatrix} 4 & 1.2 \\ 1.2 & 1 \end{pmatrix}.$$

Для новых величин

$$S = X + Y, \qquad D = X - Y$$

найдите:

1. матрицу Якоби;
2. ковариационную матрицу V_{SD};
3. коэффициент корреляции между S и D.

Задача 5. Проверка линейного приближения

Пусть

$$X \sim \mathrm{N}(1, \sigma_X^2), \qquad Y = X + aX^2.$$

1. Получите линейную оценку σ_Y.
2. Выведите точные $\mathrm{E}[Y]$ и $\mathrm{Var}(Y)$.
3. Сгенерируйте не менее 10^5 значений X и сравните выборочные моменты Y с точным и линейным результатами.
4. На сетке $0 \le a \le 1$ и $0.05 \le \sigma_X \le 1$ найдите область, в которой линейная оценка стандартного отклонения отличается от точной менее чем на 5%.

Литература

[1] Joint Committee for Guides in Metrology. Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement. 2008. doi:10.59161/JCGM100-2008E.

Глава 7

Двумерное гауссово распределение

Содержание

Аннотация

Совместное распределение двух измеряемых величин содержит больше информации, чем две ошибки, выписанные по отдельности. Ковариационная матрица задаёт масштаб и направление флуктуаций, а её обратная матрица позволяет измерить расстояние от точки до среднего. В этой главе мы построим двумерный гаусс, разберём вероятностные эллипсы и применим полную ковариацию к объединению измерений и восстановлению инвариантной массы.

7.1. Физическая задача: два показания калориметра

Рассмотрим ансамбль измерений событий с одной и той же кинематикой. Истинные энергии двух фотонов равны 60 и 40 ГэВ. Будем считать, что шкала энергии откалибрована без смещения, а отклик калориметра приближённо гауссов. Тогда средний вектор измеренных энергий равен

$$\mu = \begin{pmatrix} 60 \\ 40 \end{pmatrix} \text{ ГэВ}. \tag{7.1}$$

Дисперсии измеренных энергий и их ковариация собраны в матрицу

$$V = \begin{pmatrix} 1.44 & 0.36 \\ 0.36 & 1.00 \end{pmatrix} \text{ ГэВ}^2. \tag{7.2}$$

Вне диагонали стоит положительное число: часть ошибки у двух энергий общая. Например, обе зависят от одной калибровки шкалы энергии.

Пусть в одном событии получено

$$\mathbf{x} = \begin{pmatrix} 61.2 \\ 41.0 \end{pmatrix} \text{ ГэВ}. \tag{7.3}$$

Каждая энергия отличается от среднего ровно на своё стандартное отклонение. Этого ещё недостаточно, чтобы оценить совместное отклонение. Для него нужна квадратичная форма

$$D^2 = (\mathbf{x} - \mu)^{\mathsf{T}} V^{-1} (\mathbf{x} - \mu) = 1.54. \tag{7.4}$$

Число D^2 безразмерно. Оно учитывает и масштабы двух ошибок, и направление, в котором они коррелируют. Контур, проходящий через нашу точку, охватывает около 54% двумерного гауссова распределения. Для области с вероятностью 68.3% потребуется $D^2 \le 2.30$.

Откуда взялись эти числа и почему привычное условие $D^2 \le 1$ здесь даёт не 68.3%, а только 39.3%? Для ответа надо построить двумерный гаусс и посмотреть на геометрию его ковариационной матрицы.

7.2. От двух одномерных гауссов к двумерному

7.2.1. Независимые величины

Пусть

$$X \sim \mathcal{N}(\mu_X, \sigma_X^2), \qquad Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2) \tag{7.5}$$

и X и Y независимы. Совместная плотность равна произведению двух одномерных плотностей:

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y}\exp\left[-\frac{1}{2}\left(\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right]. \tag{7.6}$$

Введём безразмерные отклонения

$$u = \frac{x-\mu_X}{\sigma_X}, \qquad v = \frac{y-\mu_Y}{\sigma_Y}. \tag{7.7}$$

Плотность постоянна при постоянном значении

$$u^2 + v^2 = c. \tag{7.8}$$

В координатах (u, v) это окружность. В исходных координатах получается эллипс с полуосями $\sqrt{c}\,\sigma_X$ и $\sqrt{c}\,\sigma_Y$. Его оси совпадают с координатными: отклонение X ничего не сообщает об отклонении Y.

7.2.2. Корреляция поворачивает эллипс

Ковариационная матрица двух величин имеет вид

$$V = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}, \qquad |\rho| < 1. \tag{7.9}$$

Смешанный член в совместной плотности появляется вместе с корреляцией:

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2}\frac{u^2-2\rho uv+v^2}{1-\rho^2}\right]. \tag{7.10}$$

При $\rho = 0$ уравнение 7.10 переходит в произведение двух одномерных гауссов. При положительной корреляции большие значения X чаще сопровождаются большими значениями Y, поэтому эллипс наклонён вверх. Отрицательная корреляция меняет направление наклона.

Предел $|\rho| = 1$ в формулу плотности не входит. В этом случае определитель ковариационной матрицы обращается в нуль, а случайные точки лежат на прямой. Двумерной плотности относительно площади уже нет.

7.2.3. Обратная ковариационная матрица

Для матрицы 7.9

$$\det V = \sigma_X^2\sigma_Y^2(1-\rho^2), \tag{7.11}$$

а обратная матрица равна

$$V^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1/\sigma_X^2 & -\rho/(\sigma_X\sigma_Y) \\ -\rho/(\sigma_X\sigma_Y) & 1/\sigma_Y^2 \end{pmatrix}. \tag{7.12}$$

Обозначим отклонение от среднего через

$$\delta = \mathbf{x} - \mu = \begin{pmatrix} x-\mu_X \\ y-\mu_Y \end{pmatrix}. \tag{7.13}$$

Тогда показатель экспоненты можно записать одной строкой:

$$\delta^{\mathsf{T}} V^{-1} \delta = \frac{u^2 - 2\rho uv + v^2}{1 - \rho^2}. \tag{7.14}$$

Формула через ρ и матричная формула описывают одну плотность. Матрица V задаёт флуктуации, а V^{-1} оценивает отклонение с учётом этих флуктуаций.

7.2.4. Проверка коэффициента ρ

Построим коррелированные величины из двух независимых стандартных гауссов:

$$Z_1, Z_2 \sim \mathcal{N}(0,1), \qquad \operatorname{cov}(Z_1, Z_2) = 0, \tag{7.15}$$

$$X = \mu_X + \sigma_X Z_1, \qquad Y = \mu_Y + \sigma_Y \left(\rho Z_1 + \sqrt{1 - \rho^2}\, Z_2 \right). \tag{7.16}$$

Такой способ генерации пригодится и в методе Монте-Карло. Пока проверим, что параметры в нём имеют заявленный смысл.

Вывод

7.2.4.1. Средние, дисперсии и ковариация

Из уравнения 7.16 сразу следуют средние

$$\mathbb{E}[X] = \mu_X, \qquad \mathbb{E}[Y] = \mu_Y. \tag{7.17}$$

Дисперсия X равна σ_X^2. Для второй величины получаем

$$\begin{aligned} \operatorname{Var}(Y) &= \sigma_Y^2 \left[\rho^2 \operatorname{Var}(Z_1) + (1 - \rho^2) \operatorname{Var}(Z_2) \right] \\ &= \sigma_Y^2 \left[\rho^2 + (1 - \rho^2) \right] = \sigma_Y^2. \end{aligned} \tag{7.18}$$

Смешанное среднее даёт

$$\begin{aligned} \operatorname{cov}(X, Y) &= \sigma_X \sigma_Y \operatorname{cov}\left(Z_1, \rho Z_1 + \sqrt{1 - \rho^2}\, Z_2 \right) \\ &= \rho \sigma_X \sigma_Y. \end{aligned} \tag{7.19}$$

Поэтому параметр ρ в уравнении 7.10 равен коэффициенту корреляции:

$$\rho = \frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y}. \tag{7.20}$$

7.3. Многомерный гаусс и расстояние Махаланобиса

Для случайного вектора размерности d со средним μ и невырожденной ковариационной матрицей V плотность имеет вид [1]

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} \sqrt{\det V}} \exp\left[-\frac{1}{2} (\mathbf{x} - \mu)^{\mathsf{T}} V^{-1} (\mathbf{x} - \mu) \right]. \tag{7.21}$$

Квадратичная форма

$$D^2 = (\mathbf{x} - \mu)^{\mathsf{T}} V^{-1} (\mathbf{x} - \mu) \tag{7.22}$$

называется квадратом расстояния Махаланобиса. Обычное евклидово расстояние считает один ГэВ по каждой оси одинаковым отклонением. Расстояние 7.22 сначала делит отклонения на их характерные масштабы и одновременно учитывает корреляции.

7.3.1. Главные оси

Поскольку ковариационная матрица симметрична, её можно разложить как

$$V = R\Lambda R^{\mathsf{T}}, \qquad \Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_d). \tag{7.23}$$

Столбцы ортогональной матрицы R являются собственными векторами V. Перейдём к координатам

$$\mathbf{z} = \Lambda^{-1/2} R^{\mathsf{T}} (\mathbf{x} - \mu). \tag{7.24}$$

В них ковариационная матрица единична, а расстояние принимает обычный вид

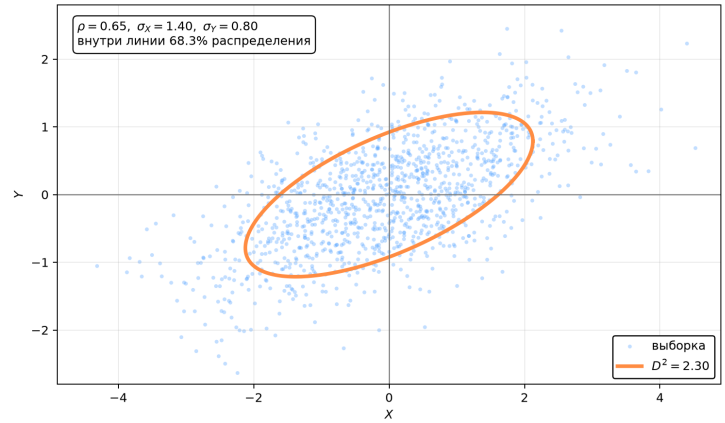
$$D^2 = \mathbf{z}^{\mathsf{T}}\mathbf{z} = \sum_{i=1}^{d} z_i^2. \tag{7.25}$$

Отсюда сразу читается геометрия контура $D^2 = c$. В двух измерениях это эллипс. Его главные оси направлены вдоль собственных векторов V, а длины полуосей равны

$$a_i = \sqrt{c\lambda_i}. \tag{7.26}$$

Коэффициент корреляции поворачивает собственные векторы относительно исходных осей. Собственные значения определяют ширины распределения в новых направлениях.

7.3.2. Интерактив: корреляция и эллипс



Интерактив. корреляция и эллипс
Меняйте ρ, σ_X и σ_Y. Точки показывают выборку из двумерного гаусса, линия — эллипс $D^2 = 2.30$.

При изменении ρ меняются сразу направление и длины осей. Матрица с теми же диагональными элементами, но другой ковариацией задаёт другое совместное распределение.

7.4. Сколько вероятности находится внутри эллипса

В одном измерении фраза «в пределах одной сигмы» понятна:

$$P(|Z| \le 1) = 0.683, \qquad Z \sim \mathcal{N}(0,1). \tag{7.27}$$

После появления второй оси требуется уточнить область. Для двумерного гаусса стандартизованные координаты независимы, поэтому

$$D^2 = Z_1^2 + Z_2^2 \sim \chi^2_2. \tag{7.28}$$

Плотность χ^2_2 особенно проста:

$$f_{\chi^2_2}(q) = \frac{1}{2} e^{-q/2}, \qquad q \ge 0. \tag{7.29}$$

Интегрируя её от нуля до c, получаем вероятность внутри эллипса:

$$P(D^2 \le c) = 1 - e^{-c/2}. \tag{7.30}$$

Для заданной вероятности P граница находится из обратной формулы

$$c = -2\ln(1-P). \tag{7.31}$$

Значение $c = 1$ в двух измерениях даёт

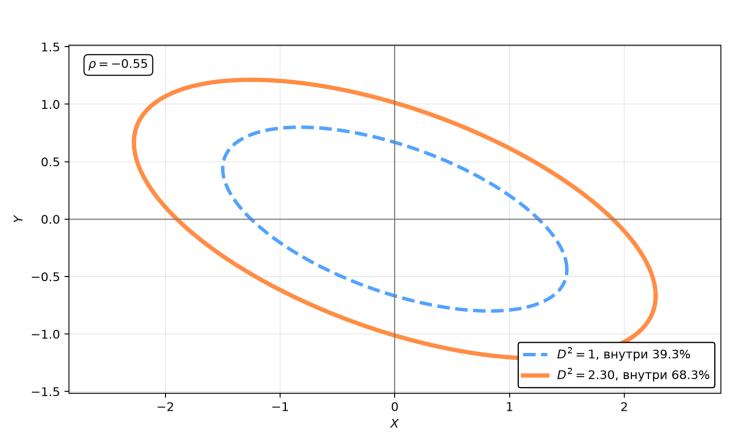
$$P(D^2 \le 1) = 1 - e^{-1/2} = 0.393. \tag{7.32}$$

В таблице 7.1 приведены уровни, которые часто встречаются на двумерных графиках.

Таблица 7.1. Вероятностные уровни двумерного гауссова распределения

Вероятность внутри контура	Граница c
39.3%	1.00
68.3%	2.30
90%	4.61
95%	5.99
99%	9.21

7.4.1. Интерактив: вероятность и размер контура



Интерактив. вероятность и размер контура
Сплошная линия ограничивает выбранную долю распределения. Пунктир показывает $D^2 = 1$, который в двух измерениях содержит 39.3% вероятности.

7.4.2. Возвращение к калориметру

Для измерения 7.3 мы нашли $D^2 = 1.54$. Уравнение 7.30 даёт

$$P(D^2 \le 1.54) = 1 - e^{-1.54/2} = 0.537. \tag{7.33}$$

Иными словами, эллипс через эту точку содержит около 54% распределения. Точка лежит внутри области 68.3%, поскольку $1.54 < 2.30$. Положительная корреляция здесь существенна: оба измерения отклонились в одну сторону, то есть вдоль вытянутой оси облака.

7.4.3. Распределение данных и пространство параметров

Эллипсы этой главы описывают случайный вектор $\mathbf{X}$ при известных μ и V. В статьях похожие контуры часто рисуют в пространстве оцениваемых параметров вокруг полученного минимума. Одинаковая квадратичная форма ещё не делает их одним объектом.

Для параметров вероятностная интерпретация зависит от того, как построен интервал: от распределения оценки в частотном подходе или от апостериорной плотности в байесовском. Значения 2.30, 4.61 и 5.99 применимы к двумерной квадратичной гауссовой задаче. Условия, при которых ими можно пользоваться для параметров, мы обсудим вместе с правдоподобием и χ^2.

7.5. Объединение измерений

Пусть n приборов измеряют одну величину μ. Результаты собраны в вектор $\mathbf{x}$, а их ковариационная матрица равна V. Если совместное распределение гауссово, значение μ, при котором полученные данные имеют наибольшую плотность, минимизирует

$$Q(\mu) = (\mathbf{x} - \mu\mathbf{1})^{\mathsf{T}} V^{-1} (\mathbf{x} - \mu\mathbf{1}). \tag{7.34}$$

Здесь $\mathbf{1}$ — вектор из единиц. Дифференцирование по μ даёт

$$\hat{\mu} = \frac{\mathbf{1}^{\mathsf{T}} V^{-1} \mathbf{x}}{\mathbf{1}^{\mathsf{T}} V^{-1} \mathbf{1}}, \qquad \operatorname{Var}(\hat{\mu}) = \frac{1}{\mathbf{1}^{\mathsf{T}} V^{-1} \mathbf{1}}. \tag{7.35}$$

Это несмещённая линейная комбинация с наименьшей дисперсией. При независимых измерениях V диагональна, и формула принимает знакомый вид

$$\hat{\mu} = \frac{\displaystyle\sum_{i=1}^{n} x_i/\sigma_i^2}{\displaystyle\sum_{i=1}^{n} 1/\sigma_i^2}, \qquad \operatorname{Var}(\hat{\mu}) = \frac{1}{\displaystyle\sum_{i=1}^{n} 1/\sigma_i^2}. \tag{7.36}$$

Для двух измерений с ковариацией c_{12} веса можно выписать явно:

$$\hat{\mu} = w_1 x_1 + w_2 x_2, \qquad w_1 = \frac{\sigma_2^2 - c_{12}}{\sigma_1^2 + \sigma_2^2 - 2c_{12}}, \qquad w_2 = \frac{\sigma_1^2 - c_{12}}{\sigma_1^2 + \sigma_2^2 - 2c_{12}}. \tag{7.37}$$

Сумма весов равна единице. Один из них может оказаться отрицательным, если корреляция велика, а точности измерений заметно различаются. Отрицательный вес помогает оценить общую флуктуацию по второму измерению и частично вычесть её из более точного результата.

7.6. Инвариантная масса двух частиц

Возьмём ещё один пример из лекции. Для двух безмассовых частиц с энергиями E_1, E_2 и углом θ между импульсами

$$m^2 = 2E_1E_2(1 - \cos\theta). \tag{7.38}$$

Вектор измеряемых величин и его ковариационная матрица равны

$$\mathbf{x} = \begin{pmatrix} E_1 \\ E_2 \\ \theta \end{pmatrix}, \qquad V_{\mathbf{x}} = \begin{pmatrix} 1.44 & 0.36 & 0 \\ 0.36 & 1.00 & 0 \\ 0 & 0 & 2.5 \cdot 10^{-5} \end{pmatrix}. \tag{7.39}$$

В энергетическом блоке числа указаны в ГэВ^2, угловая дисперсия — в рад^2. Возьмём

$$E_1 = 60.0 \text{ ГэВ}, \qquad E_2 = 40.0 \text{ ГэВ}, \qquad \theta = 0.200 \text{ рад}. \tag{7.40}$$

Градиент квадрата массы равен

$$\nabla m^2 = \begin{pmatrix} 2E_2(1 - \cos\theta) \\ 2E_1(1 - \cos\theta) \\ 2E_1E_2 \sin\theta \end{pmatrix}. \tag{7.41}$$

Вывод

7.6.0.1. Численный расчёт

Сначала найдём саму массу:

$$m^2 = 2 \cdot 60 \cdot 40(1 - \cos 0.200) = 95.680 \text{ ГэВ}^2, \qquad m = 9.782 \text{ ГэВ}. \tag{7.42}$$

В заданной точке численный градиент равен

$$\nabla m^2 = \begin{pmatrix} 1.595 \\ 2.392 \\ 953.613 \end{pmatrix}. \tag{7.43}$$

По формуле распространения ковариации из главы 6

$$\mathrm{Var}(m^2) \simeq (\nabla m^2)^{\mathsf{T}} V_{\mathbf{x}} (\nabla m^2) = 34.864\ \text{ГэВ}^4. \tag{7.44}$$

Отсюда

$$\sigma_{m^2} = 5.905\ \text{ГэВ}^2, \qquad \sigma_m \simeq \frac{\sigma_{m^2}}{2m} = 0.302\ \text{ГэВ}. \tag{7.45}$$

Результат можно записать как

$$m = (9.78 \pm 0.30)\ \text{ГэВ}. \tag{7.46}$$

Ковариация двух энергий даёт в дисперсию m^2 вклад

$$2\frac{\partial m^2}{\partial E_1}\frac{\partial m^2}{\partial E_2}\,\mathrm{cov}(E_1, E_2) = 2.75\ \text{ГэВ}^4. \tag{7.47}$$

Если этот член отбросить, получится $\sigma_m = 0.290$ ГэВ вместо 0.302 ГэВ. Та же положительная корреляция, которая наклоняла эллипс двух энергий, увеличивает ошибку массы: обе производные по энергиям положительны.

7.7. Где заканчивается гауссова геометрия

Многомерный гаусс полностью определяется средним и ковариационной матрицей. Для произвольного распределения этих двух объектов уже недостаточно. Два распределения могут иметь одинаковые μ и V, но разные хвосты и разную вероятность далёких отклонений.

Линейное распространение ковариации хорошо описывает малые флуктуации около рабочей точки. Сильная нелинейность, асимметричные входные распределения и физические границы требуют распространения полного распределения. Один из способов сделать это — сгенерировать входные величины с заданными связями и пропустить их через функцию результата.

Отдельная численная трудность возникает при $|\rho| \simeq 1$. Матрица V становится почти вырожденной, и результат обращения чувствителен к малым изменениям её элементов. Это означает, что два измерения содержат почти одну и ту же информацию. Печатать больше знаков в V^{-1} здесь бесполезно: следует проверить модель общей ошибки.

Итоги главы

- Ковариационная матрица задаёт масштабы и направления флуктуаций многомерного гаусса.
- Квадрат расстояния Махаланобиса учитывает единицы, дисперсии и корреляции.
- В двух измерениях область $D^2 \le 1$ содержит 39.3% вероятности; для 68.3% нужна граница $D^2 = 2.30$.
- Объединение измерений и распространение ошибок требуют полной ковариационной матрицы.

7.8. Задачи

Задача 1. Вероятность внутри эллипса

Для двумерного гауссова распределения

$$D^2 = (\mathbf{X} - \mu)^{\mathsf{T}} V^{-1} (\mathbf{X} - \mu) \sim \chi^2_2.$$

1. Получите плотность D^2 и покажите, что
$$P(D^2 \le c) = 1 - e^{-c/2}.$$
2. Найдите c для областей с вероятностями 68.3%, 90% и 95%.
3. Вычислите вероятность внутри контура $D^2 = 1$.
4. Сравните этот результат с одномерной вероятностью $P(|Z| \le 1)$ и объясните различие.

Задача 2. Геометрия ковариационной матрицы

Даны средний вектор и ковариационная матрица двух измеренных энергий:

$$\mu = \begin{pmatrix} 60 \\ 40 \end{pmatrix} \text{ ГэВ}, \qquad V = \begin{pmatrix} 1.44 & 0.36 \\ 0.36 & 1.00 \end{pmatrix} \text{ ГэВ}^2.$$

1. Найдите стандартные отклонения и коэффициент корреляции.
2. Вычислите собственные значения и собственные векторы V.
3. Найдите длины и направления полуосей эллипса $D^2 = 2.30$.
4. Для точки $\mathbf{x} = (61.2, 41.0)^{\mathsf{T}}$ ГэВ вычислите D^2 и определите, лежит ли она внутри области с вероятностью 68.3%.

Задача 3. Объединение двух измерений

Одна величина измерена двумя способами:

$$x_1 = 10.2 \pm 0.8, \qquad x_2 = 11.1 \pm 0.4.$$

1. Найдите взвешенное среднее и его стандартное отклонение, считая измерения независимыми.
2. Повторите расчёт для коэффициента корреляции $\rho = 0.6$.
3. Выпишите оба веса во втором случае и объясните знак каждого из них.
4. Проверьте, что ковариационная матрица положительно определена.

Задача 4. Инвариантная масса

Для двух безмассовых частиц

$$m^2 = 2E_1E_2(1 - \cos\theta).$$

Пусть $E_1 = 60$ ГэВ, $E_2 = 40$ ГэВ, $\theta = 0.200$ рад, а ковариационная матрица величин (E_1, E_2, θ) равна

$$V = \begin{pmatrix} 1.44 & 0.36 & 0 \\ 0.36 & 1.00 & 0 \\ 0 & 0 & 2.5 \cdot 10^{-5} \end{pmatrix}.$$

1. Найдите градиент m^2 и вычислите σ_{m^2}.
2. Получите σ_m линейным распространением ошибки.
3. Повторите расчёт при $\mathrm{cov}(E_1, E_2) = 0$ и сравните ответы.
4. Исследуйте предел $\theta \ll 1$: получите приближённые выражения для m и его производных.

Литература

[1] Cowan, G. *Statistical Data Analysis.* Oxford University Press. 1998.

Глава 8

Метод Монте-Карло и псевдоэксперименты

Содержание

Аннотация

Метод Монте-Карло позволяет заменить сложный расчёт большим числом случайных реализаций заданной модели. В этой главе мы разберём генерацию случайных величин, вычисление многомерных интегралов, выборку по важности и алгоритм VEGAS. Затем построим ансамбль псевдоэкспериментов и проследим цепочку от генерации физического процесса до отклика детектора и результата анализа.

8.1. Физическая задача: слабый сигнал на фоне

Эксперимент ищет редкий процесс. За время набора данных ожидается в среднем $s = 5$ сигнальных и $b = 10$ фоновых событий. Полное число зарегистрированных событий имеет распределение Пуассона:

$$N \sim \mathrm{Pois}(s + b). \tag{8.1}$$

Будем считать фон известным точно и оценим число сигнальных событий простым вычитанием:

$$\hat{s} = N - b. \tag{8.2}$$

До проведения эксперимента можно задать несколько вполне физических вопросов. Какое значение $\hat{s}$ мы в среднем получим? Каков разброс оценки? Как часто флуктуация фона приведёт к отрицательному $\hat{s}$?

Первые два ответа следуют из свойств распределения Пуассона:

$$\mathbb{E}[\hat{s}] = s, \qquad \mathrm{Var}(\hat{s}) = s + b. \tag{8.3}$$

Для $s = 5$ и $b = 10$ стандартное отклонение равно $\sqrt{15} \simeq 3.87$ события. Отрицательная оценка получается при $N < 10$. Её вероятность равна

$$P(\hat{s} < 0) = P(N \le 9 \mid s + b = 15) = \sum_{n=0}^{9} e^{-15} \frac{15^n}{n!} = 0.0699. \tag{8.4}$$

Эту задачу можно решить и численно. Сгенерируем много значений $N_k \sim \mathrm{Pois}(15)$, для каждого вычислим $\hat{s}_k = N_k - 10$ и посмотрим на полученное распределение. Его среднее, дисперсия и доля отрицательных оценок приблизятся к значениям в формулах 8.3 и 8.4.

В этой игрушечной модели аналитический ответ помещается в несколько строк. Настоящий анализ использует энергетический спектр, отклик детектора, отбор событий и мешающие параметры. Распределение результата такой процедуры обычно уже нельзя выписать одной формулой. Тогда мы повторяем эксперимент на компьютере. Это и есть метод Монте-Карло.

8.2. Что вычисляет метод Монте-Карло

Пусть случайная величина X имеет плотность $f_X(x)$, а интересующая нас величина определяется функцией g:

$$X \sim f_X(x), \qquad Y = g(X). \tag{8.5}$$

Численный алгоритм состоит из трёх действий:

1. сгенерировать независимые значения $X_1, \dots, X_N$ из распределения f_X;
2. вычислить $Y_k = g(X_k)$ для каждого события;
3. по выборке $Y_1, \dots, Y_N$ найти среднее, дисперсию, вероятность нужного события или всё распределение Y.

В короткой записи

$$X_k \sim f_X, \qquad Y_k = g(X_k), \qquad k = 1, \dots, N. \tag{8.6}$$

Один и тот же принцип используется в нескольких задачах. Интеграл можно записать как математическое ожидание,

$$\int h(x) f_X(x)\, dx = \mathbb{E}[h(X)]. \tag{8.7}$$

При распространении ошибок генерируется случайный вектор $\mathbf{X}$, после чего для каждого события вычисляется $Y = g(\mathbf{X})$. Псевдоэксперимент воспроизводит весь анализ целиком.

Метод был сформулирован как общий численный подход Метрополисом и Уламом в 1949 году [1]. Его точность определяется числом реализаций и дисперсией вычисляемой величины.

> Монте-Карло решает задачу, записанную в программе. Если в модели нет важного фона или неверно описан отклик детектора, увеличение числа событий закрепит неверный ответ с меньшей численной ошибкой. Компьютер в этом смысле безупречно послушен.

8.3. Псевдослучайные числа

Генератор псевдослучайных чисел мы уже использовали при обсуждении центральной предельной теоремы. Это детерминированный алгоритм. Его внутреннее состояние определяет всю дальнейшую последовательность

$$U_1, U_2, \dots, \qquad U_k \in [0, 1), \tag{8.8}$$

которая проходит статистические тесты как выборка из равномерного распределения. Период хорошего генератора намного превышает длину обычного расчёта.

Начальное состояние задаётся начальным числом, или *seed*. Одинаковое начальное число воспроизводит ту же последовательность. Это позволяет повторить расчёт и найти событие, на котором сломалась программа.

Для параллельной генерации нужны независимые последовательности. Если тысячу процессов запустить с одним начальным состоянием, каждый из них создаст одну и ту же тысячу событий. В файле окажется миллион строк, а независимых событий останется тысяча. Поэтому вместе с конфигурацией анализа сохраняют алгоритм генератора и правило формирования начальных состояний.

Большинство генераторов непосредственно создаёт

$$U \sim \mathrm{U}(0, 1). \tag{8.9}$$

Из равномерной величины можно получить другие распределения.

8.4. Метод обратной функции распределения

Пусть $F_X(x)$ — функция распределения требуемой величины:

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(t)\,dt. \tag{8.10}$$

Она монотонно возрастает от нуля до единицы. Если обратную функцию можно вычислить, берём $U \sim \mathrm{U}(0,1)$ и полагаем

$$\boxed{X = F_X^{-1}(U).} \tag{8.11}$$

Вывод

8.4.0.1. Почему это работает

Монотонность F_X даёт

$$\begin{aligned} P(X \le x) &= P\bigl(F_X^{-1}(U) \le x\bigr) \\ &= P(U \le F_X(x)) \\ &= F_X(x). \end{aligned} \tag{8.12}$$

В последней строке использовано $P(U \le u) = u$ для равномерной величины на $[0,1]$. Получена требуемая функция распределения. ■

8.4.1. Пример: время между событиями

Пусть события образуют пуассоновский процесс с интенсивностью λ. Время X до следующего события имеет экспоненциальную плотность

$$f_X(x) = \lambda e^{-\lambda x}, \qquad x \ge 0, \tag{8.13}$$

и функцию распределения

$$F_X(x) = 1 - e^{-\lambda x}. \tag{8.14}$$

Решая уравнение $U = F_X(X)$ относительно X, получаем

$$X = -\frac{1}{\lambda}\ln(1-U). \tag{8.15}$$

Величина $1-U$ также равномерна на $[0,1]$, поэтому в программе часто пишут

$$\boxed{X = -\frac{1}{\lambda}\ln U.} \tag{8.16}$$

8.5. Метод отбора

Обратную функцию удаётся найти не для каждого распределения. Пусть плотность $f(x)$ ограничена сверху другой плотностью $g(x)$:

$$f(x) \le M g(x), \tag{8.17}$$

причём из $g(x)$ генерировать удобно. Алгоритм отбора устроен так:

1. сгенерировать кандидата $X \sim g(x)$;
2. независимо сгенерировать $U \sim \mathrm{U}(0,1)$;

3. принять X, если
$$U \le \frac{f(X)}{Mg(X)}; \tag{8.18}$$
4. при невыполнении условия повторить попытку.

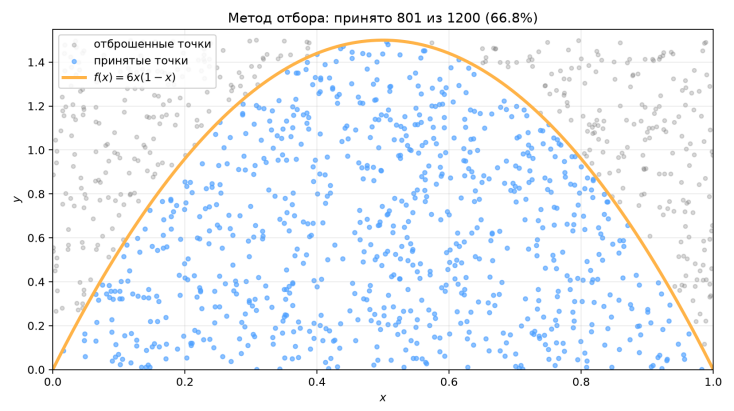
Если обе плотности нормированы, средняя доля принятых событий равна $1/M$. Чем ближе оболочка $Mg(x)$ к $f(x)$, тем меньше вычислений заканчивается отказом.

8.5.1. Пример: плотность $6x(1-x)$

Рассмотрим плотность

$$f(x) = 6x(1-x), \qquad 0 \le x \le 1. \tag{8.19}$$

Её максимум достигается при $x = 1/2$ и равен $3/2$. Равномерно генерируем точки в прямоугольнике $0 \le x \le 1$, $0 \le y \le 3/2$ и принимаем точку, если $y \le f(x)$.



Интерактив. Метод отбора для плотности $6x(1-x)$
Синие точки приняты, серые отброшены. Теоретическая доля принятых событий равна $2/3$.

8.5.2. Как выбирать способ генерации

Таблица 8.1. Основные способы генерации случайных величин

Метод	Сильная сторона	Ограничение
Обратная функция распределения	используется каждое сгенерированное число	нужна вычислимая F_X^{-1}
Метод отбора	применим к широкому классу плотностей	плохая оболочка снижает эффективность
Специальный алгоритм	может быть очень быстрым	нужна отдельная реализация
Марковская цепь	работает со сложными многомерными плотностями	значения коррелированы, сходимость требует проверки

8.6. Интегрирование методом Монте-Карло

Рассмотрим одномерный интеграл

$$I = \int_a^b h(x)\,dx. \tag{8.20}$$

Если $U \sim \mathrm{U}(a,b)$, то

$$\mathbb{E}[h(U)] = \frac{1}{b-a}\int_a^b h(x)\,dx. \tag{8.21}$$

Математическое ожидание заменим выборочным средним:

$$\boxed{\hat{I}_N = \frac{b-a}{N}\sum_{k=1}^{N} h(U_k).} \tag{8.22}$$

8.6.1. Статистическая ошибка интеграла

Вывод

8.6.1.1. Дисперсия оценки $\hat{I}_N$

Обозначим

$$\mathrm{Var}[h(U)] = \sigma_h^2. \tag{8.23}$$

Для независимых точек дисперсии слагаемых складываются:

$$\begin{aligned}\mathrm{Var}(\hat{I}_N) &= \frac{(b-a)^2}{N^2}\,\mathrm{Var}\left(\sum_{k=1}^{N} h(U_k)\right)\\ &= \frac{(b-a)^2\sigma_h^2}{N}.\end{aligned} \tag{8.24}$$

Следовательно,

$$\boxed{\sigma_{\hat{I}_N} = \frac{(b-a)\sigma_h}{\sqrt{N}}.} \tag{8.25}$$

■

Чтобы уменьшить ошибку в десять раз, число точек придётся увеличить в сто раз. Это медленная сходимость. Её важное достоинство проявляется в большой размерности: показатель $N^{-1/2}$ не превращается в $N^{-1/d}$. Коэффициент перед ним, конечно, зависит от конкретной функции и выбранного распределения точек.

8.7. Проверка на числе π

Равномерно сгенерируем точки в единичном квадрате:

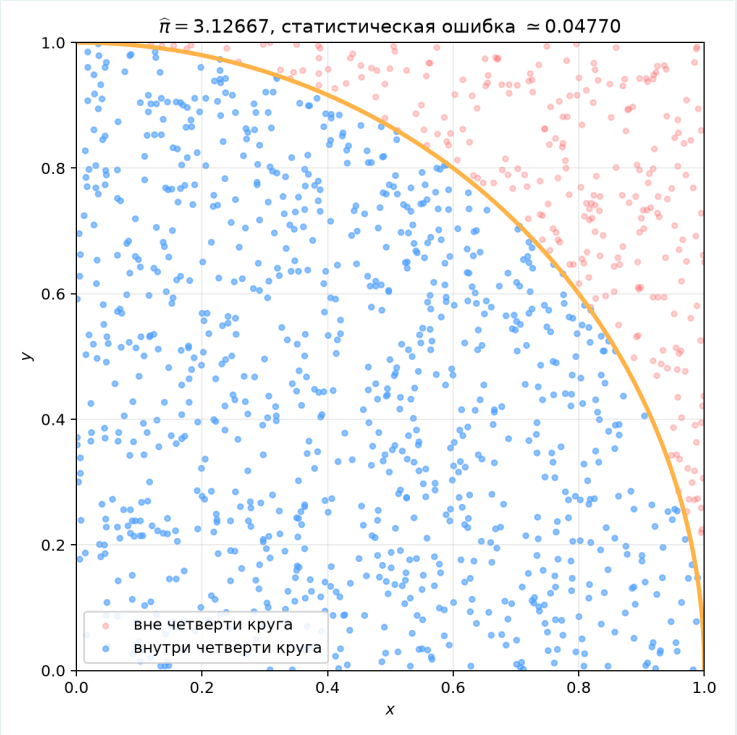
$$(X,Y) \sim \mathrm{U}([0,1]^2). \tag{8.26}$$

Точка лежит внутри четверти единичного круга при $X^2 + Y^2 \le 1$. Поэтому

$$P(X^2+Y^2 \le 1) = \frac{\pi}{4}, \tag{8.27}$$

и доля попавших внутрь точек даёт оценку

$$\widehat{\pi} = 4\frac{N_{\text{in}}}{N}. \tag{8.28}$$



Интерактив. Оценка числа π методом Монте-Карло
Доля синих точек внутри четверти круга оценивает $\pi/4$.

Если единственная цель состоит в вычислении числа π, калькулятор справится лучше. Этот пример ценен другим: точный ответ известен, геометрия видна на рисунке, а ошибка подчиняется закону $1/\sqrt{N}$ из формулы 8.25. Такие задачи нужны для проверки реализации метода.

8.8. Высокая размерность

Регулярная сетка с m узлами по каждой координате в d измерениях содержит

$$N_{\text{grid}} = m^d \tag{8.29}$$

точек. Сто узлов по каждой из шести координат дают 10^{12} вычислений функции. Сетка ещё довольно грубая, а вычислительная задача уже нет.

Регулярная сетка

5 точек по каждой координате
в 8-мерном пространстве

N = 390 625

Интерактив. Рост числа узлов регулярной сетки
При пяти узлах по каждой из восьми координат требуется $5^8 = 390\,625$ вычислений.

Регулярная сетка

12 точек по каждой координате
в 8-мерном пространстве

N = 429 981 696

Рис. 8.1. *При двенадцати узлах по каждой из восьми координат число вычислений возрастает до* $12^8 = 429\,981\,696$

В одном измерении хорошая квадратурная формула обычно выигрывает у случайного бросания точек. В большой размерности сетка платит за все комбинации координат. Метод Монте-Карло использует ровно столько точек, сколько мы сгенерировали. Его преимущество следует проверять на конкретном интеграле.

8.9. Выборка по важности

Пусть требуется вычислить

$$I = \int h(x) f(x)\, dx. \tag{8.30}$$

Равномерная генерация не учитывает форму интегранда. Если основной вклад сосредоточен в малой области, большая часть вычислений почти ничего не добавит

к результату. Выберем плотность $q(x)$, положительную везде, где $h(x)f(x) \neq 0$, и перепишем интеграл:

$$\boxed{I = \mathbb{E}_q\left[h(X)\frac{f(X)}{q(X)}\right].} \tag{8.31}$$

Точка $X_k \sim q(x)$ получает вес

$$w_k = \frac{f(X_k)}{q(X_k)}, \tag{8.32}$$

а оценка интеграла равна

$$\hat{I}_N = \frac{1}{N}\sum_{k=1}^{N} w_k h(X_k). \tag{8.33}$$

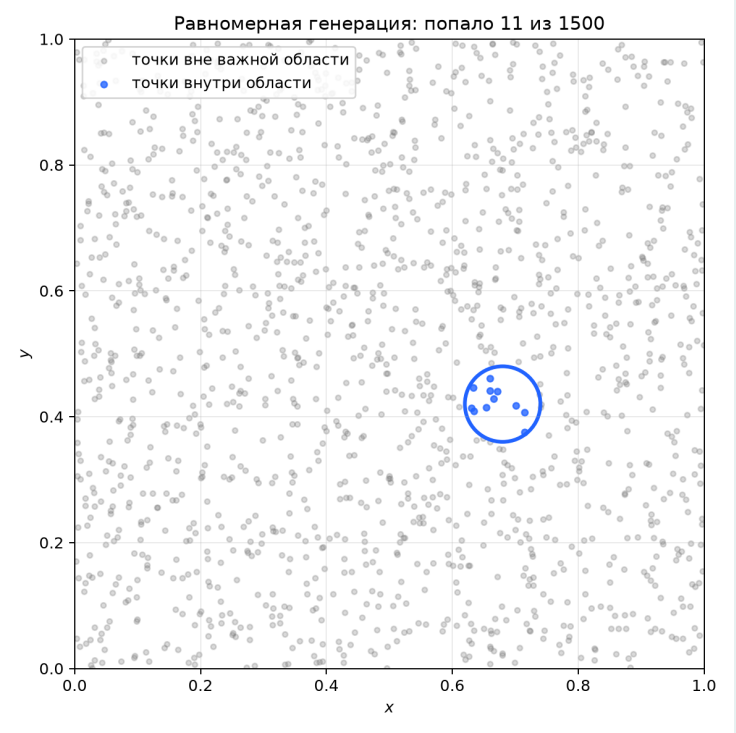
Хорошая плотность $q(x)$ часто приводит нас в область большого вклада и даёт веса одного порядка. Если редкое событие получает огромный вес, оно может определить весь ответ и его дисперсию. Число строк в файле тогда плохо характеризует статистическую точность. Для положительных весов полезно вычислять эффективный размер выборки

$$N_{\text{eff}} = \frac{\left(\sum_k w_k\right)^2}{\sum_k w_k^2}. \tag{8.34}$$

При одинаковых весах $N_{\text{eff}} = N$. Сильно различающиеся веса дают $N_{\text{eff}} \ll N$.

8.9.1. Редкая область

Следующий интерактив показывает одну и ту же малую область при двух способах генерации. Равномерная выборка покрывает весь квадрат. Выборка по важности сосредоточивает точки вблизи области, которая даёт основной вклад. При численном интегрировании изменение плотности обязательно компенсируется весом из формулы 8.32.



Интерактив. Равномерная выборка и выборка по важности
При равномерной генерации малая доля точек попадает в важную область.

8.10. VEGAS: адаптивная выборка по важности

Для ручного выбора $q(\mathbf{x})$ надо заранее знать важную область. Алгоритм VEGAS строит приближение к такой плотности по первым итерациям расчёта [2]. В каждой координате он сгущает сетку там, где интегранд вносит большой вклад. Полученная плотность имеет приближённо разделяющийся вид

$$q(\mathbf{x}) = \prod_{i=1}^{d} q_i(x_i). \tag{8.35}$$

VEGAS сохраняет асимптотический закон $1/\sqrt{N}$. Выигрыш получается за счёт меньшей дисперсии. Разделяющаяся плотность хорошо описывает пики, положение которых видно по отдельным координатам. Узкий диагональный гребень или несколько далеко разнесённых пиков представляют более трудную задачу.

8.10.1. Контрольный восьмимерный интеграл

Рассмотрим произведение нормированных гауссовых плотностей в единичном гиперкубе:

$$f(\mathbf{x}) = \prod_{i=1}^{8} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_i - \mu_i)^2}{2\sigma^2}\right], \qquad 0 \le x_i \le 1, \tag{8.36}$$

где $\sigma = 0.10$, а центры равны

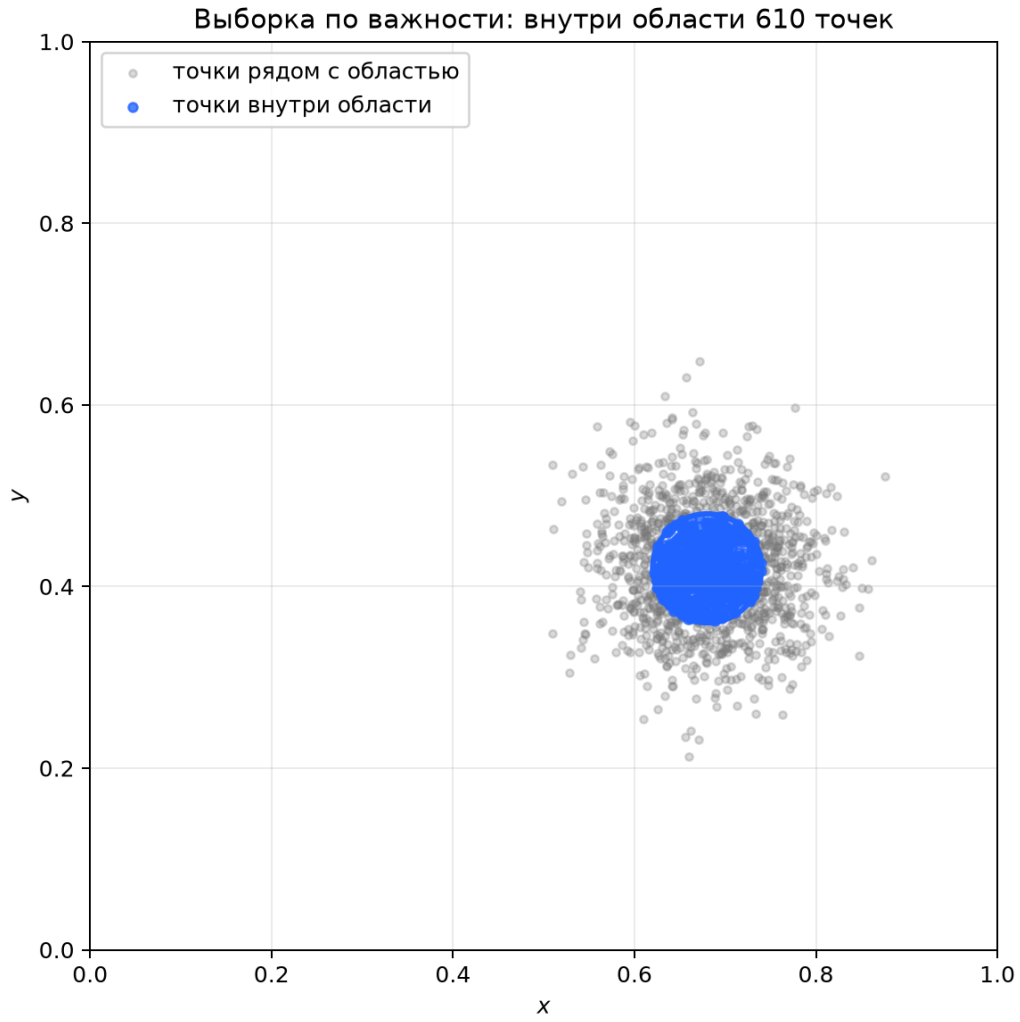


Рис. 8.2. *Выборка по важности сосредоточивает точки вблизи области, которая даёт основной вклад в интеграл*

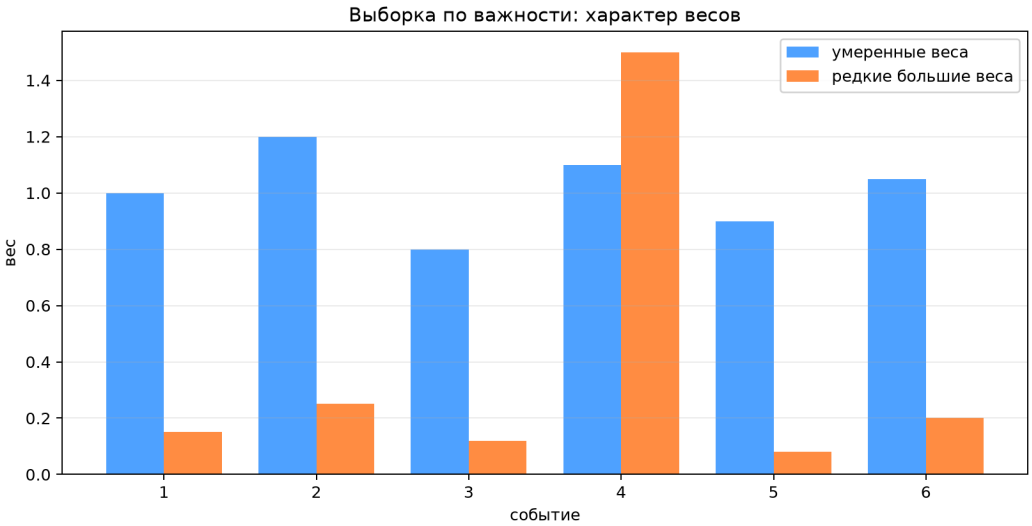


Рис. 8.3. *Умеренные веса дают много сопоставимых вкладов; редкие события с очень большими весами увеличивают дисперсию оценки*

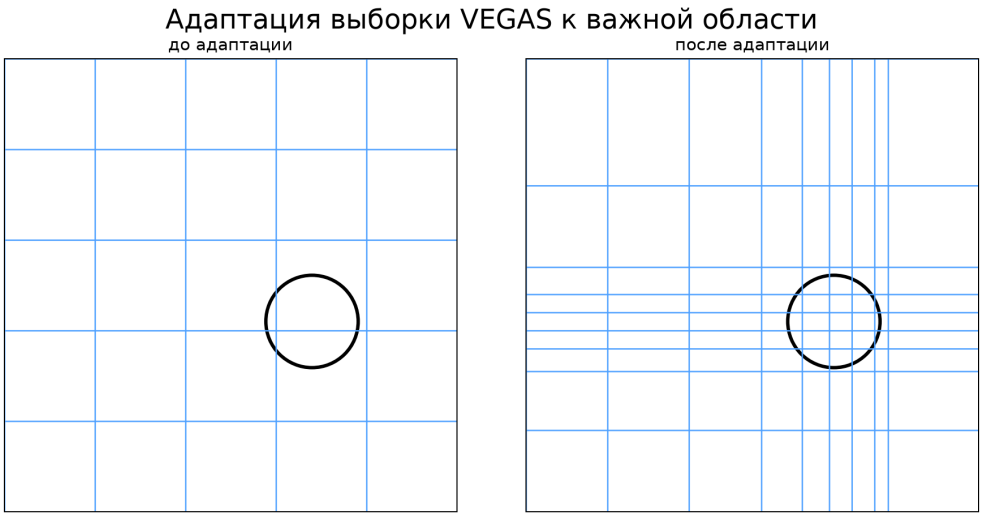


Рис. 8.4. *После нескольких итераций VEGAS располагает больше интервалов в области большого вклада*

$$\mu = (0.63, 0.37, 0.58, 0.42, 0.69, 0.31, 0.54, 0.46). \tag{8.37}$$

Точный интеграл находится как произведение восьми одномерных гауссовых интегралов:

$$I_{\mathrm{exact}} = 0.9978196359. \tag{8.38}$$

Простой Монте-Карло с одним и тем же начальным числом даёт результаты из таблицы 8.2.

Таблица 8.2. Равномерный метод Монте-Карло для восьмимерного интеграла

N	$\hat{I}_N$
10^4	0.946 ± 0.48
10^5	0.864 ± 0.13
10^6	1.001 ± 0.061

Миллион точек всё ещё даёт ошибку около шести процентов: почти вся равномерная выборка проходит мимо узкого пика. Тензорная квадратура Гаусса—Лежандра сходится, но число вызовов функции растёт как m^8.

Таблица 8.3. Тензорная квадратура для восьмимерного интеграла

Узлов на координату m	Всего узлов m^8	Результат
3	6 561	0.589
5	390 625	0.722
7	5 764 801	0.963

Числа в обеих таблицах получены скриптом `shared/notebooks/mc_vegas_demo.py` с начальным состоянием 12345. После восьми адаптационных итераций по 20 000 точек и десяти рабочих итераций по 50 000 точек VEGAS даёт

$$\hat{I}_{\text{VEGAS}} = 0.99787 \pm 0.00053. \tag{8.39}$$

Этот пример специально удобен для VEGAS: интегранд раскладывается в произведение функций отдельных координат. Для нового интеграла результат проверяют независимыми запусками, увеличением числа точек и сравнением с более грубым методом. Очень узкий пик алгоритм может не встретить на первых итерациях, а значит, и не научиться генерировать точки в его окрестности.

8.11. Псевдоэксперименты

Псевдоэксперимент — одна искусственная реализация измерения при заданной модели. В простом случае достаточно сгенерировать одно число. В полном анализе последовательно выполняются генерация физического процесса, моделирование детектора, реконструкция, отбор и та же статистическая процедура, которая применяется к данным.

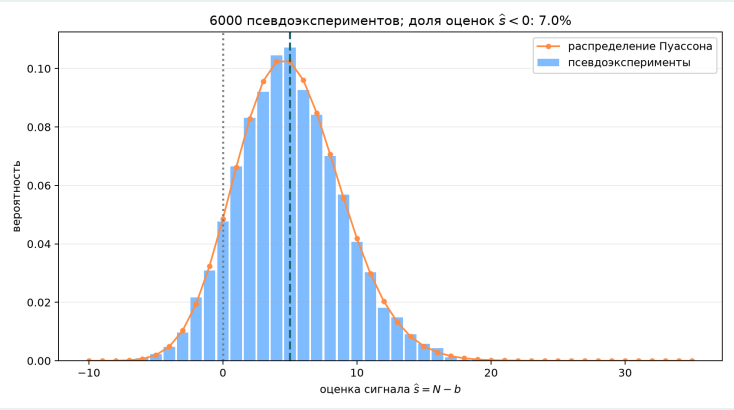
Ансамбль псевдоэкспериментов позволяет найти:

- смещение и дисперсию оценки параметра;
- распределение статистики теста;
- покрытие доверительного интервала;
- ожидаемую чувствительность эксперимента;
- долю неудачных или нестабильных подгонок;
- изменение результата при систематических вариациях модели.

Каждый такой вывод относится к модели и значениям параметров, при которых был сгенерирован ансамбль.

8.11.1. Возвращаемся к счётному эксперименту

Для модели из формулы 8.1 один псевдоэксперимент состоит из генерации $N \sim \text{Pois}(s + b)$ и вычисления $\hat{s} = N - b$. Повторив эти действия несколько тысяч раз, получим распределение оценки.



Интерактив. Ансамбль счётных экспериментов
При $s = 5$ и $b = 10$ среднее распределения равно пяти, а около семи процентов псевдоэкспериментов дают $\hat{s} < 0$.

Отрицательное $\hat{s}$ не означает отрицательного числа физических событий. Так проявляется флуктуация оценки $N - b$. Статистическая процедура должна уметь

работать и с такими реализациями; удаление их из ансамбля изменит среднее и нарушит свойства метода.

Реальный эксперимент даёт одну точку из показанного распределения. Увеличение числа псевдоэкспериментов точнее определяет это распределение, но не уменьшает разброс результата одного реального измерения.

8.12. Моделирование полного эксперимента

В физике частиц и нейтрино один псевдоэксперимент может опираться на длинную цепочку моделирования:

1. генератор задаёт процесс, типы частиц и их кинематику;
2. моделируется прохождение частиц через вещество, потери энергии и вторичные частицы;
3. вычисляется отклик сенсоров и электроники;
4. программа реконструкции восстанавливает энергию, координаты и тип события;
5. применяются триггер, критерии качества и отбор анализа.

Каждый этап содержит приближения. Их проверяют по контрольным данным, а оставшиеся расхождения входят в систематическую неопределённость.

Интерактив. Моделирование оптических фотонов в Daya Bay и JUNO
В ролике Саймона Блита показана работа Opticks: распространение оптических фотонов рассчитывается на графическом процессоре для геометрий детекторов Daya Bay и JUNO. Описание проекта опубликовано на странице автора [3].

8.12.1. Матрица отклика

Пусть истинный спектр разбит на бины. Элемент матрицы отклика

$$R_{ji} = P(\text{событие восстановлено в бине } j \mid \text{истинный бин } i) \tag{8.40}$$

оценивается по моделированию. Если истинный спектр равен $\mathbf{t}$, то ожидаемый восстановленный спектр имеет вид

$$\boxed{\mathbf{r} = R\mathbf{t}.} \tag{8.41}$$

Диагональные элементы описывают правильную реконструкцию бина, внедиагональные — миграции. Сумма элементов столбца может быть меньше единицы: часть событий теряется из-за геометрического аксептанса, порогов и отбора.

8.12.2. Эффективность отбора

Если из N_{gen} сгенерированных событий прошло отбор N_{sel}, оценка эффективности равна

$$\hat{\varepsilon} = \frac{N_{\text{sel}}}{N_{\text{gen}}}. \tag{8.42}$$

Для независимых событий с единичными весами

$$N_{\text{sel}} \sim \text{Bin}(N_{\text{gen}}, \varepsilon), \tag{8.43}$$

поэтому стандартную ошибку обычно оценивают как

$$\boxed{\sigma_{\hat{\varepsilon}} \simeq \sqrt{\frac{\hat{\varepsilon}(1-\hat{\varepsilon})}{N_{\text{gen}}}}.} \tag{8.44}$$

У этой формулы хорошо видны границы применимости. Если все события прошли отбор, она даёт $\hat{\varepsilon} = 1$ и нулевую ошибку. Из конечной выборки следует лишь, что в ней не встретился ни один отказ. Следующее событие пройти отбор не обязано. Вблизи физических границ $0 \le \varepsilon \le 1$ нужны интервальные методы; они рассматриваются в следующих главах.

При наличии весов биномиальная модель уже неприменима в таком виде. Тогда проверяют суммы весов, суммы квадратов весов и эффективный размер выборки из формулы 8.34.

8.13. Как проверять расчёт

Расчёт Монте-Карло начинают с задач, для которых известен ответ. Проверяют нормировку, средние, дисперсии, корреляции и предельные значения параметров. Результаты независимых запусков должны быть совместимы, а численная ошибка — убывать с ожидаемой скоростью при увеличении статистики.

Если из псевдоэксперимента восстанавливается параметр θ, на ансамбле с известным истинным значением вычисляют смещение

$$\text{bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta \tag{8.45}$$

и дисперсию

$$\text{Var}(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2\right]. \tag{8.46}$$

Оценка не обязана быть точно несмещённой при конечной статистике. Её смещение должно быть измерено и сопоставлено с требуемой точностью анализа.

> Здесь участвуют два разных разброса. Физическая статистическая неопределённость описывает возможные результаты одного реального эксперимента. Численная ошибка Монте-Карло показывает, насколько точно конечная симуляция оценивает это распределение. Большее число псевдоэкспериментов уменьшает вторую ошибку. Первая определяется объёмом реальных данных.

Для взвешенной выборки отдельно проверяют распределение весов и N_{eff}. Для адаптивного интегрирования сравнивают итерации и независимые начальные состояния. Для полного моделирования тестируют каждый этап цепочки и сопоставляют контрольные распределения с данными.

Вероятностная модель, конфигурация, версии программ и начальные состояния генераторов являются частью результата. Без них численный ответ трудно воспроизвести, а иногда невозможно даже понять, что именно было вычислено.

8.14. Когда нужен другой метод

Для одномерного интеграла с гладкой известной функцией квадратура обычно быстрее простого Монте-Карло. Редкие области требуют выборки по важности или специализированной генерации. Для сложных многомерных распределений параметров используют методы Монте-Карло по марковским цепям и вложенную выборку. Полное моделирование детектора иногда заменяют быстрой параметризацией или суррогатной моделью.

Выбор метода определяется задачей и стоимостью одного вычисления. Результат Монте-Карло считается надёжным после проверки реализации, сходимости и модели, по которой он получен.

Итоги главы

- Метод Монте-Карло строит численный ансамбль реализаций заданной модели.
- Равномерные псевдослучайные числа преобразуются в нужное распределение обратной функцией, методом отбора или специальным алгоритмом.
- Ошибка простого интегрирования убывает как $1/\sqrt{N}$; в большой размерности это часто выгоднее регулярной сетки.
- Выборка по важности и VEGAS уменьшают дисперсию, направляя вычисления в области большого вклада.
- Псевдоэксперименты позволяют проверить всю статистическую процедуру от генерации данных до оценки параметра.
- Статистическая неопределённость реального эксперимента и численная ошибка конечной симуляции являются разными величинами.

8.15. Задачи

Задача 1. Обратная функция распределения

Пусть

$$U \sim \mathrm{U}(0,1).$$

1. Покажите, что
$$X = -\frac{1}{\lambda}\ln(1-U)$$
имеет экспоненциальное распределение с параметром λ.
2. Сгенерируйте выборку X и сравните гистограмму с аналитической плотностью.
3. Проверьте выборочные среднее и дисперсию.

Задача 2. Метод отбора

Требуется генерировать плотность

$$f(x) = 6x(1-x), \qquad 0 \le x \le 1.$$

1. Найдите максимум $f(x)$.
2. Постройте алгоритм отбора из равномерного распределения.
3. Вычислите теоретическую эффективность алгоритма.
4. Проверьте её численно.

Задача 3. Интегрирование методом Монте-Карло

Вычислите

$$I = \int_0^1 \frac{4}{1+x^2}\,dx = \pi$$

методом Монте-Карло.

1. Запишите оценку $\hat{I}_N$.
2. Оцените её статистическую ошибку по выборке.
3. Исследуйте зависимость ошибки от N.
4. Покажите численно, что ошибка убывает приблизительно как $N^{-1/2}$.

Задача 4. Редкое событие и выборка по важности

Требуется оценить вероятность

$$p = P(Z > 4), \qquad Z \sim \mathrm{N}(0,1).$$

1. Оцените p наивной генерацией из стандартного гаусса.
2. Объясните, почему для устойчивой оценки требуется очень большая выборка.
3. Предложите распределение $q(z)$, чаще генерирующее область $z > 4$.
4. Запишите соответствующий вес $w(z) = f(z)/q(z)$.
5. Сравните дисперсии двух оценок.

Задача 5. Псевдоэксперименты для счётного эксперимента

Пусть

$$N \sim \mathrm{Pois}(s+b), \qquad \hat{s} = N - b,$$

где фон b известен точно.

1. Найдите аналитические $\mathbb{E}[\hat{s}]$ и $\mathrm{Var}(\hat{s})$.
2. Постройте ансамбль псевдоэкспериментов.
3. Сравните выборочные среднее и дисперсию с аналитическими значениями.
4. Оцените вероятность получить $\hat{s} < 0$.

Задача 6. Нелинейное распространение ошибки

Пусть коррелированные величины имеют распределение

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} 10 \\ 5 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right].$$

Для отношения

$$R = \frac{X}{Y}$$

сравните:

1. линейную оценку ошибки;
2. стандартное отклонение выборки Монте-Карло;
3. центральный квантильный интервал $[q_{0.16}, q_{0.84}]$.

Исследуйте зависимость результатов от σ_Y и ρ.

Задача 7. Эффективность отбора

Из $N_{\text{gen}} = 10\,000$ сгенерированных событий отбор прошло $N_{\text{sel}} = 730$.

1. Оцените эффективность отбора.
2. Оцените её биномиальную статистическую ошибку.
3. Сколько событий требуется сгенерировать, чтобы относительная ошибка эффективности была меньше 1%?
4. Обсудите, как изменится задача при наличии весов событий.

Задача 8. Эффективный размер выборки

Для взвешенной выборки определено

$$N_{\text{eff}} = \frac{(\sum_k w_k)^2}{\sum_k w_k^2}.$$

1. Покажите, что при одинаковых весах $N_{\text{eff}} = N$.
2. Найдите N_{eff} для весов
$$(1, 1, 1, 1, 6).$$
3. Сравните результат с фактическим числом событий.
4. Объясните, почему большая вариация весов ухудшает статистическую точность.

Литература

[1] Metropolis, N. и Ulam, S. The Monte Carlo Method. *Journal of the American Statistical Association*, **44**, 335–341. 1949. doi:10.1080/01621459.1949.10483310.
[2] Lepage, G. P. A New Algorithm for Adaptive Multidimensional Integration. *Journal of Computational Physics*, **27**, 192–203. 1978. doi:10.1016/0021-9991(78)90004-9.
[3] Blyth, S. C. Opticks: GPU Optical Photon Simulation. https://juno.ihep.ac.cn/~blyth/.

Глава 9

Правдоподобие и оценки максимального правдоподобия

Содержание

Аннотация

В этой главе мы перейдём от вероятностной модели данных к оценке её параметров. Начнём с калибровки эффективности детектора, введём функцию правдоподобия и найдём её максимум. Затем тот же метод применим к пуассоновскому счёту и гауссову распределению. Последний пример покажет, что оценка максимального правдоподобия может быть смещённой.

9.1. Физическая задача: эффективность регистрации

Перед исследуемым детектором установлен другой прибор, который отмечает прохождение заряженной частицы. За время калибровки он зарегистрировал $N = 20$ частиц. Исследуемый детектор сработал для $k = 7$ из них. Требуется оценить его эффективность ε. Ошибками первого прибора и случайными совпадениями пока пренебрежём.

Каждая частица даёт один из двух результатов: детектор сработал или не сработал. Если результаты независимы, а эффективность во время калибровки постоянна, число зарегистрированных частиц имеет биномиальное распределение:

$$K \sim \mathrm{Bin}(N, \varepsilon). \tag{9.1}$$

Первая оценка напрашивается сама:

$$\hat{\varepsilon} = \frac{k}{N} = \frac{7}{20} = 0.35. \tag{9.2}$$

Почему именно эта оценка? Что означает её шляпка? Как изменится вывод, если детектор сработает семь раз из двадцати или семьдесят раз из двухсот? И что делать, если число зарегистрированных событий складывается из сигнала и известного фона? Для ответа запишем функцию правдоподобия.

9.2. Оценка и её значение

Пусть данные описываются распределением

$$X \sim f(x \mid \theta), \tag{9.3}$$

где θ — параметр или набор параметров модели. Это может быть эффективность детектора, среднее число событий, масса частицы, сечение процесса или коэффициент корреляции.

В частотном подходе θ фиксирован, но неизвестен. До эксперимента случайны данные. **Оценкой** параметра называется правило, которое ставит данным в соответствие число или вектор чисел:

$$\hat{\theta} = T(X_1, \dots, X_N). \tag{9.4}$$

После подстановки конкретной выборки получается значение оценки:

$$\hat{\theta}_{\text{data}} = T(x_1, \dots, x_N). \tag{9.5}$$

Это различие понадобится дальше. В формуле 9.4 оценка является случайной величиной и меняется от опыта к опыту. Число в формуле 9.5 уже получено из данных.

9.2.1. Один параметр — разные оценки

Пусть $X_1, \dots, X_N$ имеют гауссово распределение со средним μ. Оценивать μ можно выборочным средним, медианой или, например, полусуммой первого и последнего элементов выборки:

$$\hat{\mu}_1 = \frac{1}{N}\sum_{i=1}^{N} X_i, \qquad \hat{\mu}_2 = \operatorname{median}(X_1, \dots, X_N), \qquad \hat{\mu}_3 = \frac{X_1 + X_N}{2}. \tag{9.6}$$

Все три правила дают оценку одного параметра. При гауссовом распределении они несмещённые, но разброс у них различен. Третья оценка, кроме того, выбрасывает почти всю собранную информацию. Условие задачи само по себе ещё не выбирает оценку. Нужен принцип, по которому этот выбор будет сделан.

Точечная оценка даёт одно значение $\hat{\theta}$. Интервальная оценка задаёт область допустимых значений параметра. В этой главе мы строим точечные оценки. Интервалы потребуют отдельной статистической процедуры, к которой мы перейдём позже.

9.3. Функция правдоподобия

При заданном θ функция $f(x \mid \theta)$ описывает распределение возможных данных. После эксперимента данные $\mathbf{x}$ фиксированы, а параметр остаётся неизвестным. Рассмотрим ту же функцию как функцию параметра:

$$\boxed{L(\theta; \mathbf{x}) = f(\mathbf{x} \mid \theta)}. \tag{9.7}$$

Она называется **функцией правдоподобия**. Вертикальная черта в $f(\mathbf{x} \mid \theta)$ напоминает, что распределение данных задаётся при фиксированных параметрах. Точка с запятой в $L(\theta; \mathbf{x})$ отделяет аргумент функции от уже полученных данных.

Таблица 9.1. Вероятностная модель и функция правдоподобия используют одну математическую функцию, но отвечают на разные вопросы.

Функция	Фиксировано	Что меняется	Вопрос
$f(\mathbf{x} \mid \theta)$	θ	$\mathbf{x}$	Как распределены результаты повторных опытов?
$L(\theta; \mathbf{x})$	$\mathbf{x}$	θ	Какие параметры лучше согласуются с этими данными?

Правдоподобие не является распределением вероятности параметра. Для него нет условия нормировки по θ. Более того, множитель, который не зависит от θ, можно отбросить: положение максимума и отношения значений L от этого не изменятся.

9.3.1. Независимая выборка

Для независимых измерений совместное правдоподобие равно произведению:

$$L(\theta) = \prod_{i=1}^{N} f_i(x_i \mid \theta). \tag{9.8}$$

Индекс у f_i допускает разные распределения или разные ошибки отдельных измерений. Для одинаково распределённых величин все f_i совпадают.

Произведение многих малых чисел быстро достигает машинного нуля. Поэтому в вычислениях используют логарифм правдоподобия:

$$\ell(\theta) = \ln L(\theta) = \sum_{i=1}^{N} \ln f_i(x_i \mid \theta). \tag{9.9}$$

Логарифм монотонен, поэтому L и ℓ достигают максимума при одних и тех же параметрах. Независимые наборы данных добавляют к общей функции по одному слагаемому:

$$\ell_{\text{total}}(\theta) = \ell_1(\theta) + \ell_2(\theta) + \cdots. \tag{9.10}$$

В больших экспериментах результаты разных каналов и периодов набора данных объединяются сложением соответствующих логарифмов.

9.4. Метод максимального правдоподобия

Оценка максимального правдоподобия выбирает параметры, при которых полученные данные имеют наибольшее правдоподобие [1, 2]:

$$\boxed{\hat{\theta}_{\text{ММП}} = \arg\max_{\theta} L(\theta) = \arg\max_{\theta} \ell(\theta)}. \tag{9.11}$$

Если максимум находится внутри допустимой области, а функция гладкая, то для одного параметра выполняются условия

$$\left.\frac{d\ell}{d\theta}\right|_{\theta=\hat{\theta}} = 0, \qquad \left.\frac{d^2\ell}{d\theta^2}\right|_{\theta=\hat{\theta}} < 0. \tag{9.12}$$

Слово «внутри» здесь существенно. Если параметр ограничен, максимум может оказаться на границе, где первая производная не равна нулю. Такой случай мы увидим уже в биномиальной задаче.

БИОГРАФИЯ

Рональд Эйлмер Фишер

1890–1962

Рональд Фишер — британский статистик и генетик. В работах 1920-х годов он придал современную форму теории статистического оценивания, ввёл понятия достаточной статистики и информации Фишера и систематически развил метод максимального правдоподобия.

Работая на Ротамстедской опытной станции, Фишер создал методы планирования и анализа эксперимента, включая дисперсионный анализ. В генетике он связал менделевское наследование с естественным отбором и стал одним из основателей популяционной генетики. Его имя встретится ещё раз при обсуждении информации, которую данные содержат о параметре.

Источник: MacTutor; фото неизвестный фотограф, 1913; общественное достояние

9.5. Биномиальная модель

Вернёмся к калибровке детектора. При N независимых частицах и эффективности ε вероятность зарегистрировать ровно k частиц равна

$$P(K = k \mid \varepsilon) = \binom{N}{k} \varepsilon^k (1-\varepsilon)^{N-k}. \tag{9.13}$$

После того как N и k получены, эта формула рассматривается как функция ε. Комбинаторный множитель от эффективности не зависит, поэтому

$$L(\varepsilon) \propto \varepsilon^k (1-\varepsilon)^{N-k}. \tag{9.14}$$

9.5.1. Оценка эффективности

Сначала рассмотрим $0 < k < N$, когда максимум лежит внутри интервала $0 < \varepsilon < 1$. Логарифм правдоподобия с точностью до постоянного слагаемого имеет вид

$$\ell(\varepsilon) = k \ln \varepsilon + (N-k) \ln(1-\varepsilon) + \text{const}. \tag{9.15}$$

Его производная равна

$$\frac{d\ell}{d\varepsilon} = \frac{k}{\varepsilon} - \frac{N-k}{1-\varepsilon}. \tag{9.16}$$

Условие максимума даёт

$$k(1-\varepsilon) = \varepsilon(N-k),$$

откуда

$$\boxed{\hat{\varepsilon}_{\text{ММП}} = \frac{k}{N}}. \tag{9.17}$$

Так получено значение оценки для конкретных данных. До эксперимента вместо k следует писать случайную величину K:

$$\hat{\varepsilon} = \frac{K}{N}. \tag{9.18}$$

Из свойств биномиального распределения следуют среднее и дисперсия оценки:

$$\mathbb{E}[\hat{\varepsilon}] = \varepsilon, \qquad \operatorname{Var}(\hat{\varepsilon}) = \frac{\varepsilon(1-\varepsilon)}{N}. \tag{9.19}$$

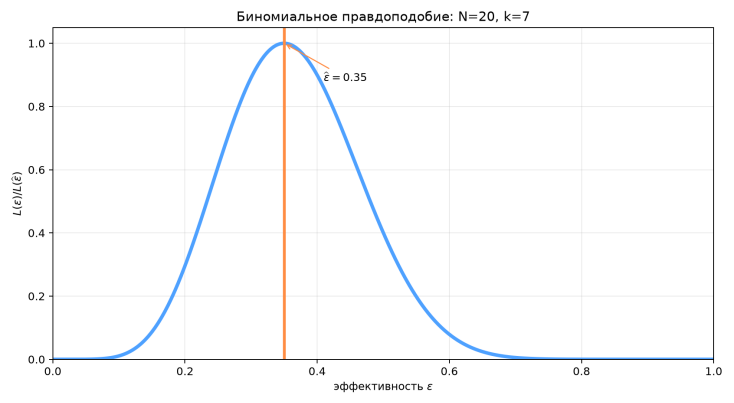
Значит, в этой задаче оценка максимального правдоподобия несмещённая, а её разброс убывает как $1/\sqrt{N}$. Неизвестную ε в формуле для стандартного отклонения часто заменяют её оценкой:

$$\hat{\sigma}_{\hat{\varepsilon}} = \sqrt{\frac{\hat{\varepsilon}(1-\hat{\varepsilon})}{N}}. \tag{9.20}$$

Для нашей калибровки $\hat{\varepsilon} = 0.35$ и $\hat{\sigma}_{\hat{\varepsilon}} \simeq 0.107$. Последнее число пока не следует читать как готовый доверительный интервал. Особенно плохо такая интерпретация работает возле $\varepsilon = 0$ и $\varepsilon = 1$.

Если $k = 0$, функция правдоподобия монотонно убывает и максимум находится при $\hat{\varepsilon} = 0$. При $k = N$ она монотонно возрастает и максимум равен $\hat{\varepsilon} = 1$. В обоих случаях максимум лежит на физической границе, а уравнение 9.16 для внутренней точки применять нельзя.

9.5.2. Интерактив: форма биномиального правдоподобия



Интерактив. Биномиальное правдоподобие
Меняйте число частиц N и число срабатываний k. Вертикальная линия показывает оценку $\hat{\varepsilon} = k/N$. При увеличении выборки максимум становится уже.

На рисунке 9.1 отношение k/N сохранено, но число испытаний увеличено в пять раз. Положение максимума не изменилось, а значения эффективности вдали от 0.35 теперь подавлены гораздо сильнее.

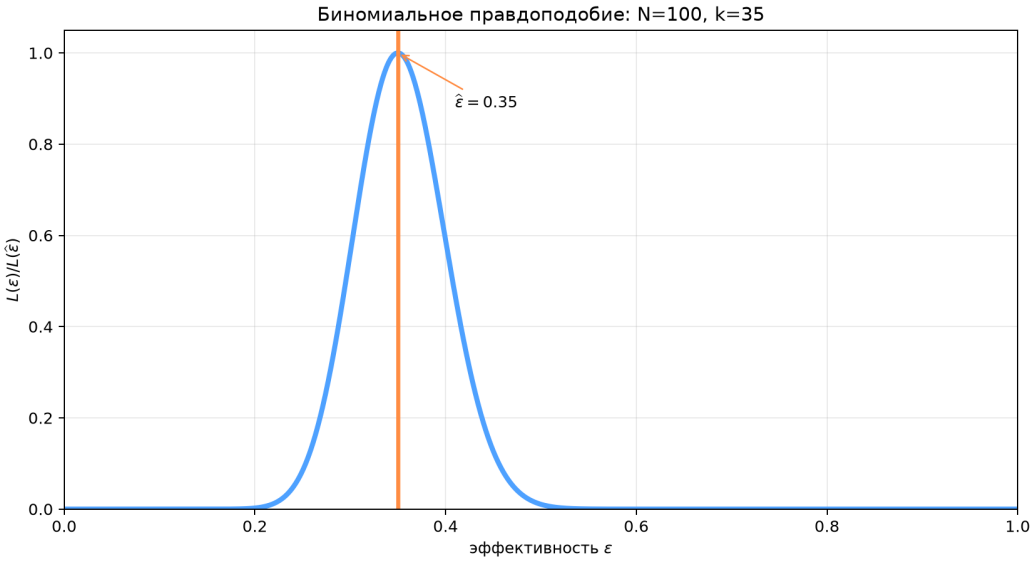


Рис. 9.1. *Относительное правдоподобие эффективности при $N = 100$ и $k = 35$. Вертикальная линия соответствует $\hat{\varepsilon} = 0.35$.*

Биномиальная модель предполагает, что каждая отмеченная первым прибором частица действительно прошла через исследуемый детектор, а вероятность его срабатывания оставалась постоянной. Фоновые совпадения, зависимость эффективности от энергии и неоднородность детектора добавят в правдоподобие новые параметры. Одна дробь k/N такую калибровку уже не описывает.

9.6. Пуассоновский счёт

Пусть N — число событий, зарегистрированных за фиксированное время, а μ — ожидаемое число событий:

$$N \sim \mathrm{Pois}(\mu). \tag{9.21}$$

Если получено n событий, правдоподобие равно

$$L(\mu) = \frac{\mu^n e^{-\mu}}{n!}, \qquad \mu \geq 0. \tag{9.22}$$

Для $n > 0$

$$\ell(\mu) = n \ln \mu - \mu - \ln n!, \qquad \frac{d\ell}{d\mu} = \frac{n}{\mu} - 1. \tag{9.23}$$

Следовательно,

$$\boxed{\hat{\mu}_{\text{ММП}} = n}. \tag{9.24}$$

При $n = 0$ максимум находится на границе $\mu = 0$, и ответ остаётся тем же. Поскольку до эксперимента $\hat{\mu} = N$, а $\mathbb{E}[N] = \mu$, эта оценка несмещённая.

9.6.1. Известный фон

Теперь пусть среднее число событий складывается из неизвестного сигнала s и известного фона b:

$$N \sim \mathrm{Pois}(s + b), \qquad s \geq 0. \tag{9.25}$$

Если временно разрешить s принимать любые значения при условии $s + b \geq 0$, максимум находится при

$$\hat{s}_{\text{unconstrained}} = n - b. \tag{9.26}$$

Однако отрицательное среднее число сигнальных событий не имеет физического смысла. С учётом ограничения $s \geq 0$ получаем

$$\boxed{\hat{s}_{\text{ММП}} = \max(0, n - b)}. \tag{9.27}$$

При $n < b$ производная логарифма правдоподобия в разрешённой области не обращается в нуль. Максимум находится в точке $s = 0$. Физическая граница меняет статистическую процедуру, и обычное гауссово приближение возле максимума в этой ситуации требует отдельной проверки.

9.7. Среднее гауссова распределения

Пусть независимые измерения имеют одинаковую известную ошибку σ:

$$X_i \sim \mathrm{N}(\mu, \sigma^2), \qquad i = 1, \ldots, N. \tag{9.28}$$

Требуется оценить μ. Логарифм правдоподобия с точностью до постоянного слагаемого равен

$$\ell(\mu) = -\frac{1}{2\sigma^2} \sum_{i=1}^{N} (x_i - \mu)^2 + \text{const}. \tag{9.29}$$

Максимизация ℓ совпадает с минимизацией суммы квадратов отклонений. Из условия

$$\frac{d\ell}{d\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{N}(x_i - \mu) = 0 \tag{9.30}$$

получаем

$$\boxed{\hat{\mu}_{\text{ММП}} = \frac{1}{N}\sum_{i=1}^{N} x_i = \overline{x}}. \tag{9.31}$$

Метод максимального правдоподобия выбрал из трёх оценок 9.6 выборочное среднее. Связь между гауссовым правдоподобием и методом наименьших квадратов будет подробно использована в следующих главах.

Если ошибки отдельных измерений σ_i различаются, в сумме появляется вес $1/\sigma_i^2$. Максимум правдоподобия тогда даёт взвешенное среднее, полученное в формуле 7.36.

9.8. Неизвестны среднее и дисперсия

Пусть теперь требуется одновременно оценить μ и σ^2. Полный логарифм правдоподобия имеет вид

$$\ell(\mu, \sigma^2) = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i - \mu)^2. \tag{9.32}$$

Нормировочный множитель гауссова распределения теперь зависит от σ^2, поэтому отбрасывать первое слагаемое нельзя. Максимизация по двум параметрам даёт

$$\boxed{\hat{\mu}_{\text{ММП}} = \overline{x}, \qquad \widehat{\sigma^2}_{\text{ММП}} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}. \tag{9.33}$$

Оценка среднего несмещённая. Для дисперсии получается другой результат:

$$\mathbb{E}\left[\widehat{\sigma^2}_{\text{ММП}}\right] = \frac{N-1}{N}\sigma^2. \tag{9.34}$$

Делитель N появился из условия максимума правдоподобия. Несмещённая оценка дисперсии использует делитель $N - 1$:

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2, \qquad \mathbb{E}[s^2] = \sigma^2. \tag{9.35}$$

Здесь оценка максимального правдоподобия и несмещённая оценка расходятся. Метод максимального правдоподобия задаёт способ построения оценки. Несмещённость — отдельное свойство этой оценки. Для $\widehat{\sigma^2}_{\text{ММП}}$ смещение исчезает при $N \to \infty$, но при конечном N оно существует.

9.9. Что уже можно прочитать по функции правдоподобия

Положение максимума даёт точечную оценку. Ширина максимума показывает, как быстро ухудшается согласие модели с данными при смещении параметра. С ростом выборки максимум обычно сужается, как в биномиальном примере. Если максимум прижат к границе, форма функции становится несимметричной.

Эти наблюдения пока остаются качественными. Чтобы превратить ширину правдоподобия в ошибку оценки или интервал, потребуется определить статистическую процедуру и проверить её свойства. Этому посвящена следующая глава.

Итоги главы

- Оценка является функцией случайных данных; после подстановки выборки она принимает конкретное значение.
- Правдоподобие сравнивает значения параметров при фиксированных данных и не является распределением вероятности параметра.
- Для независимых измерений правдоподобия перемножаются, а их логарифмы складываются.
- Оценка максимального правдоподобия соответствует максимуму L в допустимой области параметров.
- Метод максимального правдоподобия не гарантирует несмещённость: для гауссовой дисперсии оценка с делителем N смещена при конечной выборке.

9.10. Задачи

Задача 1. Вероятность и правдоподобие

Одно наблюдение x получено из экспоненциального распределения

$$f(x \mid \lambda) = \lambda e^{-\lambda x}, \qquad x \geq 0.$$

1. Нарисуйте $f(x \mid \lambda)$ как функцию x при нескольких фиксированных λ.
2. После наблюдения $x = x_{\text{obs}}$ запишите $L(\lambda; x_{\text{obs}})$.
3. Вычислите $\int_0^\infty L(\lambda; x_{\text{obs}})\, d\lambda$. Почему этот интеграл не является условием нормировки вероятности параметра?
4. Найдите $\hat{\lambda}_{\text{ММП}}$.
5. Объясните словами различие между плотностью вероятности и функцией правдоподобия.

Задача 2. Биномиальная оценка

В N независимых испытаниях наблюдалось k успехов.

1. Запишите биномиальное правдоподобие $L(p)$.
2. Найдите $\hat{p}_{\text{ММП}}$.
3. Вычислите $\mathbb{E}[\hat{p}]$ и $\mathrm{Var}(\hat{p})$.
4. Покажите, что оценка несмещённая и состоятельная.
5. Исследуйте форму $L(p)$ для случаев $k = 0$, $k = N/2$ и $k = N$.

Задача 3. Пуассоновский счёт с известным фоном

Пусть

$$N \sim \mathrm{Pois}(s + b),$$

где $s \geq 0$ — неизвестный сигнал, а фон b известен точно.

1. Запишите $L(s)$ и $\ell(s)$.
2. Найдите неограниченную оценку максимального правдоподобия сигнала.
3. Найдите оценку максимального правдоподобия при физическом ограничении $s \geq 0$.
4. Объясните, что происходит при $n < b$.

Задача 4. Гауссово среднее

Пусть

$$X_i \sim \mathrm{N}(\mu, \sigma^2),$$

где σ известна.

1. Запишите функцию правдоподобия независимой выборки.
2. Покажите, что $\hat{\mu}_{\mathrm{ММП}} = \langle X \rangle$.
3. Найдите $\mathrm{Var}(\hat{\mu}_{\mathrm{ММП}})$.

Задача 5. Взвешенное среднее

Имеются независимые измерения

$$X_i \sim \mathrm{N}(\mu, \sigma_i^2),$$

где σ_i известны и различаются.

1. Запишите совместный логарифм функции правдоподобия.
2. Найдите оценку максимального правдоподобия параметра μ.
3. Покажите, что результат является взвешенным средним с весами $w_i = 1/\sigma_i^2$.
4. Найдите дисперсию оценки.
5. Объясните, почему измерения с меньшей ошибкой получают больший вес.

Литература

[1] Fisher, R. A. On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society of London. Series A*, **222**, 309–368. 1922. doi:10.1098/rsta.1922.0009.
[2] Cowan, G. *Statistical Data Analysis.* Oxford University Press. 1998.

Глава 10

Свойства оценок и профильное правдоподобие

Содержание

Аннотация

В прошлой главе мы научились находить максимум правдоподобия. Теперь разберём, какими свойствами обладает полученная оценка и как определить её ошибку. Для этого введём смещение, средний квадрат ошибки и информацию Фишера, исследуем форму правдоподобия и научимся учитывать мешающие параметры. Примерами послужат измерение времени жизни частицы и подгонка спектра реакторных антинейтрино.

10.1. Физическая задача: время жизни частицы

Детектор зарегистрировал $N = 20$ распадов. Для каждого события измерено время от рождения частицы до её распада. Сумма этих времён оказалась равна 40 мкс. Как оценить среднее время жизни τ?

В простейшей модели времена независимы и имеют экспоненциальное распределение. Фоном, разрешением прибора и потерями событий пренебрежём. Предполагается, что регистрация одинаково эффективна при любом времени распада. Тогда оценка максимального правдоподобия равна среднему по выборке:

$$\hat{\tau} = \frac{1}{N}\sum_{i=1}^{N} t_i = 2.0 \text{ мкс}. \tag{10.1}$$

А какая у неё ошибка? Если использовать свойства экспоненциального распределения, стандартное отклонение оценки получится равным $\tau/\sqrt{N}$. Истинного τ мы пока не знаем. Подставляя вместо него оценку, получим

$$\hat{\sigma}_{\hat{\tau}} = \frac{\hat{\tau}}{\sqrt{N}} \simeq 0.45 \text{ мкс}. \tag{10.2}$$

Двадцать событий дали нам время жизни с относительной статистической ошибкой около 22%. Можно ли при том же числе событий придумать оценку поточнее? Насколько обоснована симметричная запись 2.0 ± 0.45 мкс? И что изменится, когда в модели появится неизвестный фон?

Для ответа надо посмотреть, как оценка ведёт себя при повторении эксперимента. Все ожидания и дисперсии ниже относятся к ансамблю возможных данных при фиксированном истинном параметре.

10.2. Свойства оценок

10.2.1. Смещение, дисперсия и средний квадрат ошибки

Пусть мы много раз повторили один и тот же эксперимент и каждый раз получили оценку $\hat{\theta}$. Среднее этих оценок может отличаться от истинного значения θ. Разность называется **смещением**:

$$b(\theta) = \operatorname{bias}(\hat{\theta}) = \mathbb{E}_\theta[\hat{\theta}] - \theta. \tag{10.3}$$

Если $b(\theta) = 0$ для всех допустимых θ, оценка несмещённая. Это свойство среднего по ансамблю. Отдельное измерение, конечно, не обязано попасть в истинное значение. Разброс оценок около их собственного среднего описывается дисперсией $\operatorname{Var}_\theta(\hat{\theta})$.

Чтобы учесть и смещение, и разброс, введём **средний квадрат ошибки**, или MSE (*mean squared error*):

$$\mathrm{MSE}_\theta(\hat\theta) = \mathbb{E}_\theta[(\hat\theta - \theta)^2]. \tag{10.4}$$

Здесь отклонение отсчитывается от истинного параметра. Добавим и вычтем $m = \mathbb{E}_\theta[\hat\theta]$:

$$\hat\theta - \theta = (\hat\theta - m) + (m - \theta).$$

После возведения в квадрат смешанный член исчезает при усреднении, поскольку $\mathbb{E}_\theta[\hat\theta - m] = 0$. Остаётся

$$\boxed{\mathrm{MSE}_\theta(\hat\theta) = \mathrm{Var}_\theta(\hat\theta) + b^2(\theta).} \tag{10.5}$$

Получается, что устранение смещения само по себе ещё не гарантирует меньшей ошибки. Посмотрим на пример, который уже встретился нам в прошлой главе.

10.2.2. Два делителя в оценке дисперсии

Для независимой гауссовой выборки размера $N \geq 2$ с неизвестными μ и σ^2 имеем две оценки:

$$\widehat{\sigma^2}_{\text{ММП}} = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2, \qquad s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})^2. \tag{10.6}$$

Первая получается максимизацией правдоподобия. Вторая несмещённая. Их средние равны

$$\mathbb{E}[\widehat{\sigma^2}_{\text{ММП}}] = \frac{N-1}{N}\sigma^2, \qquad \mathbb{E}[s^2] = \sigma^2. \tag{10.7}$$

Но у s^2 больше разброс: каждый результат первой оценки умножен на $N/(N-1)$. Для гауссовой выборки точный расчёт даёт

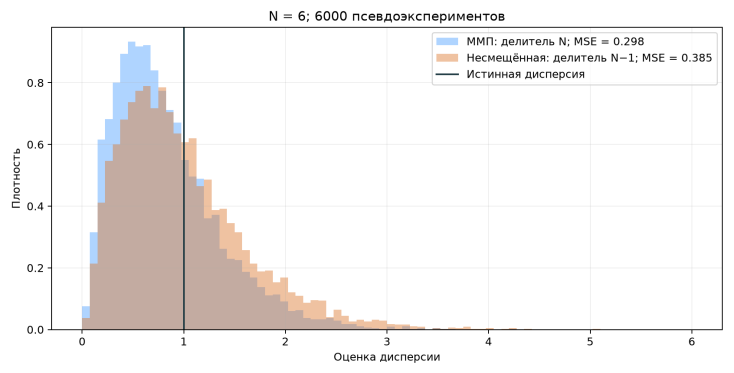
$$\mathrm{Var}(\widehat{\sigma^2}_{\text{ММП}}) = \frac{2(N-1)}{N^2}\sigma^4, \qquad \mathrm{Var}(s^2) = \frac{2}{N-1}\sigma^4. \tag{10.8}$$

Подставим эти выражения в формулу 10.5:

$$\mathrm{MSE}(\widehat{\sigma^2}_{\text{ММП}}) = \frac{2N-1}{N^2}\sigma^4, \qquad \mathrm{MSE}(s^2) = \frac{2}{N-1}\sigma^4. \tag{10.9}$$

При $N = 6$ коэффициенты перед σ^4 равны примерно 0.306 и 0.400. Смещённая оценка в этом сравнении выигрывает по среднему квадрату ошибки. Это не делает её лучшей для любой задачи: мы выбрали конкретный критерий и посчитали его в конкретной модели.

В апплете истинная дисперсия равна единице. Для каждой из 6000 гауссовых выборок вычисляются обе оценки. Сравните положения гистограмм, их ширины и средние квадраты ошибок. Затем увеличьте N.



Интерактив. Две оценки дисперсии
Распределения оценок дисперсии для 6000 гауссовых выборок размера $N = 6$ при $\sigma^2 = 1$. Делитель $N - 1$ устраняет смещение и одновременно увеличивает разброс оценки.

10.2.3. Состоятельность

Что должно происходить с оценкой, когда данных становится всё больше? Мы хотим, чтобы вероятность заметно промахнуться мимо истинного параметра стремилась к нулю. Это свойство называется **состоятельностью**:

$$P_\theta\big(|\hat\theta_N - \theta| > \varepsilon\big) \xrightarrow[N\to\infty]{} 0 \quad \text{при любом } \varepsilon > 0. \tag{10.10}$$

Короткая запись этого условия — $\hat\theta_N \xrightarrow{P} \theta$. Например, закон больших чисел обеспечивает состоятельность выборочного среднего при существующем конечном математическом ожидании.

Удобный достаточный признак состоятельности — одновременно исчезающие смещение и дисперсия. Действительно,

$$P_\theta\big(|\hat\theta_N - \theta| > \varepsilon\big) \le \frac{\mathrm{MSE}_\theta(\hat\theta_N)}{\varepsilon^2}. \tag{10.11}$$

Если правая часть стремится к нулю, выполнено определение 10.10. Обратного вывода без дополнительных условий нет: редкие, но очень большие отклонения могут мешать сходимости моментов.

10.2.4. Сравнение трёх оценок гауссова среднего

В прошлой главе мы предложили три правила оценки μ: среднее по всей выборке, медиану и полусумму первого и последнего измерений. Для независимых $X_i \sim \mathcal{N}(\mu, \sigma^2)$ получаем

$$\begin{aligned} \mathrm{Var}(\overline{X}) &= \frac{\sigma^2}{N}, \\ \mathrm{Var}(\text{median } X_i) &\simeq \frac{\pi}{2}\frac{\sigma^2}{N}, \\ \mathrm{Var}\left(\frac{X_1 + X_N}{2}\right) &= \frac{\sigma^2}{2}, \qquad N \ge 2. \end{aligned} \tag{10.12}$$

Первая и третья формулы точные; формула для медианы асимптотическая, при большом N. Среднее и медиана состоятельны. У третьей оценки распределение

вообще не меняется при добавлении данных: два использованных измерения остаются двумя измерениями. Несмещённость здесь есть, состоятельности нет.

В чистой гауссовой модели среднее точнее медианы. Зато медиана слабее реагирует на отдельные выбросы. Если в данные попадают события из другого процесса, сначала надо разобраться с моделью выборки. Вывод об оптимальности среднего был получен для гауссовой модели без такой примеси.

10.3. Информация Фишера и граница Крамера–Рао

Можно ли найти нижнюю границу дисперсии оценки, ещё не выбрав саму оценку? В регулярных моделях это позволяет сделать информация Фишера [1].

Обозначим логарифм правдоподобия всей выборки через $\ell(\theta)$ и введём его производную — **скор-функцию** (*score*):

$$U(\theta) = \frac{\partial \ell(\theta)}{\partial \theta}. \tag{10.13}$$

До подстановки данных $U(\theta)$ является случайной величиной. Для информации Фишера существуют две равносильные записи:

$$\boxed{I(\theta) = \mathbb{E}_\theta[U^2(\theta)] = -\mathbb{E}_\theta\left[\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right].} \tag{10.14}$$

В каких условиях они равносильны? Область возможных данных не должна зависеть от параметра; плотность должна быть достаточно гладкой, чтобы производные можно было переносить под интеграл. В дальнейшем также предполагаем $0 < I(\theta) < \infty$. Посмотрим, где используются эти условия.

10.3.1. Две формы информации Фишера

Для одного наблюдения положим $u = \partial_\theta \ln f(x \mid \theta)$. Дифференцируя условие нормировки, получаем

$$0 = \frac{\partial}{\partial \theta} \int f(x \mid \theta)\, dx = \int f(x \mid \theta) u \, dx = \mathbb{E}_\theta[u]. \tag{10.15}$$

Ещё одно дифференцирование даёт

$$\begin{aligned} 0 &= \int \frac{\partial^2 f}{\partial \theta^2}\, dx \\ &= \int f\left[u^2 + \frac{\partial^2 \ln f}{\partial \theta^2}\right] dx. \end{aligned}$$

Отсюда следует формула 10.14 для одного наблюдения. Для независимых наблюдений логарифмы правдоподобия и их производные складываются. Средние скор-функций равны нулю, поэтому смешанные члены в квадрате суммы исчезают:

$$I_N(\theta) = \sum_{i=1}^{N} I_i(\theta). \tag{10.16}$$

Для одинаково распределённых наблюдений $I_N = N I_1$. Это и есть количественная мера того, как растёт информация при накоплении статистики.

10.3.2. Граница Крамера–Рао

Для несмещённой оценки $T(\mathbf{X})$ дифференцируем равенство $\mathbb{E}_\theta[T] = \theta$. Если производную снова можно перенести под интеграл, получаем

$$1 = \mathbb{E}_\theta[TU] = \mathbb{E}_\theta[(T-\theta)U].$$

По неравенству Коши–Буняковского квадрат среднего произведения не больше произведения средних квадратов. Следовательно,

$$1 \le \operatorname{Var}_\theta(T)\, I(\theta), \qquad \boxed{\operatorname{Var}_\theta(T) \ge \frac{1}{I(\theta)}.} \tag{10.17}$$

Это **граница Крамера–Рао**. Несмещённую оценку, достигающую границы, называют эффективной. Достижимость не гарантируется для произвольной модели. Смещённые оценки также не обязаны подчиняться границе именно в таком виде: в её выводе мы использовали $\mathbb{E}_\theta[T] = \theta$.

10.3.3. Пример: гауссово среднее

Для одного гауссова измерения с известной σ скор-функция равна

$$u_\mu(X) = \frac{X-\mu}{\sigma^2}, \qquad I_1(\mu) = \frac{\mathbb{E}[(X-\mu)^2]}{\sigma^4} = \frac{1}{\sigma^2}. \tag{10.18}$$

Для N независимых измерений $I_N = N/\sigma^2$. Значит, дисперсия любой несмещённой оценки μ не меньше σ^2/N. Выборочное среднее имеет ровно эту дисперсию. В принятой модели мы уже получили эффективную оценку.

10.4. Форма правдоподобия около максимума

Информация Фишера содержит усреднение по возможным данным. После эксперимента мы располагаем одной выборкой и одной функцией $\ell(\theta)$. Её кривизну можно вычислить непосредственно.

Пусть максимум находится внутри допустимой области, а ℓ гладкая. Разложим её около $\hat\theta$. Первая производная в максимуме равна нулю:

$$\ell(\theta) \simeq \ell(\hat\theta) - \frac{1}{2} J(\hat\theta)(\theta-\hat\theta)^2, \qquad J(\hat\theta) = -\ell''(\hat\theta). \tag{10.19}$$

Величина $J(\hat\theta)$ определяется конкретной выборкой. При регулярной модели и достаточно большом N обратная кривизна даёт оценку дисперсии ММП:

$$\hat\sigma^2_{\hat\theta} \simeq \frac{1}{J(\hat\theta)}. \tag{10.20}$$

Узкий максимум соответствует малой ошибке. Но само наличие максимума ещё не означает, что он хорошо описывается параболой. Асимметрию, границы и другие максимумы надо проверять по всей функции правдоподобия.

10.4.1. Отношение правдоподобий

Разделим $L(\theta)$ на его максимальное значение:

$$\lambda(\theta) = \frac{L(\theta)}{L(\hat\theta)}. \tag{10.21}$$

Теперь нормировка сократилась, а максимум равен единице. Для вычислений удобнее пользоваться логарифмом:

$$\boxed{q(\theta) = -2\ln\lambda(\theta) = -2[\ell(\theta) - \ell(\hat\theta)].} \tag{10.22}$$

В максимуме $q = 0$, в остальных точках $q \geq 0$. В квадратичном приближении

$$q(\theta) \simeq \frac{(\theta - \hat\theta)^2}{\hat\sigma_{\hat\theta}^{\,2}}. \tag{10.23}$$

Поэтому точки $q = 1$ отстоят от оценки на одну стандартную ошибку. На графике ℓ им соответствует спуск от максимума на $1/2$.

Для гауссовой выборки с известной σ получаем точное равенство

$$q(\mu) = \frac{(\mu - \bar x)^2}{\sigma^2/N}. \tag{10.24}$$

Здесь интервал $q(\mu) \leq 1$ имеет покрытие около 68.3%: при повторении эксперимента такая процедура накрывает истинное μ с этой частотой. Для других моделей тот же порог требует обоснования.

Теорема Уилкса даёт его в пределе большой выборки: при истинном значении одного проверяемого параметра статистика q асимптотически имеет распределение χ^2 с одной степенью свободы [2]. Нужны идентифицируемая гладкая модель, невырожденная информация и истинная точка внутри области параметров. У физической границы или при слабой чувствительности эти условия могут нарушаться. Частотное покрытие тогда проверяют отдельно, например на псевдоэкспериментах.

10.5. Возвращаемся к времени жизни

В нашей модели $\tau > 0$ — среднее время жизни, а плотность времени распада равна

$$f(t \mid \tau) = \frac{1}{\tau} e^{-t/\tau}, \qquad t \geq 0. \tag{10.25}$$

Для N независимых времён

$$L(\tau) = \tau^{-N} \exp\!\left(-\frac{1}{\tau}\sum_i t_i\right), \qquad \ell(\tau) = -N\ln\tau - \frac{1}{\tau}\sum_i t_i. \tag{10.26}$$

Как обычно, в логарифме опущена константа, не зависящая от параметра. Условие максимума имеет вид

$$\frac{\partial \ell}{\partial \tau} = -\frac{N}{\tau} + \frac{\sum_i t_i}{\tau^2} = 0, \qquad \hat\tau = \bar t. \tag{10.27}$$

Так мы получили формулу, с которой начали главу. У экспоненциальной величины $\mathbb{E}[T_i] = \tau$ и $\mathrm{Var}(T_i) = \tau^2$, поэтому

$$\mathbb{E}[\hat\tau] = \tau, \qquad \mathrm{Var}(\hat\tau) = \frac{\tau^2}{N}. \tag{10.28}$$

Оценка несмещённая и состоятельная. Информация Фишера равна

$$I_N(\tau) = -\mathbb{E}_\tau\!\left[\frac{N}{\tau^2} - \frac{2\sum_i T_i}{\tau^3}\right] = \frac{N}{\tau^2}. \tag{10.29}$$

Дисперсия 10.28 достигает границы Крамера–Рао при любом N. Среди несмещённых оценок сделать её меньше в этой модели нельзя.

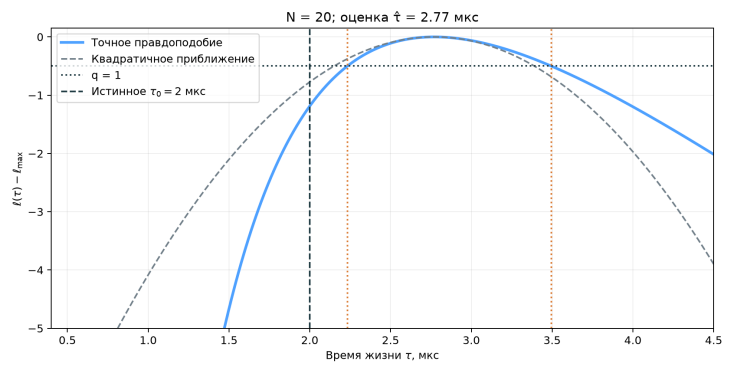
А вот распределение оценки при малом N ещё не гауссово: сумма экспоненциальных времён имеет гамма-распределение. Эффективность оценки и гауссова форма её распределения — разные свойства. Само правдоподобие также отличается от гауссова. Для него получаем точное выражение:

$$q(\tau) = 2N\left[\ln\frac{\tau}{\hat{\tau}} + \frac{\hat{\tau}}{\tau} - 1\right]. \tag{10.30}$$

Оно асимметрично относительно $\hat{\tau}$. Для наших $N = 20$ и $\hat{\tau} = 2.0$ мкс решения уравнения $q(\tau) = 1$ равны примерно 1.61 и 2.52 мкс. Расстояния до оценки различаются: 0.39 вниз и 0.52 мкс вверх. Симметричная ошибка 0.45 мкс передаёт местную кривизну, но теряет эту асимметрию. Само решение $q = 1$ на конечной выборке ещё не гарантирует точного покрытия 68.3%.

10.5.1. Численный эксперимент

В апплете генерируются времена распадов при заданном истинном τ_0. Двигайте пробное τ и найдите максимум. Затем включите границы $q = 1$. Повторите опыт несколько раз при малом N, а после увеличьте выборку. Максимум меняется от опыта к опыту; масштаб этого разброса задаётся формулой 10.28.



Интерактив. Оценка времени жизни
Одна выборка из $N = 20$ распадов при $\tau_0 = 2$ мкс. Показаны разность $\ell(\tau) - \ell_{\max}$, её квадратичное приближение и границы $q = 1$. Все данные в этом примере сгенерированы.

На рисунке 10.1 сравниваются две выборки. При большем N характерная ширина правдоподобия уменьшается. Это не обещает, что в каждом следующем опыте оценка будет ближе к истине: статистические флуктуации сохраняются.

10.6. Малые выборки и физические границы

Ещё один пример — пуассоновский счёт. После регистрации n событий оценка ожидаемого числа равна $\hat{\mu} = n$. Из пуассоновского правдоподобия получаем

$$q(\mu) = 2\left[\mu - n + n\ln\frac{n}{\mu}\right]. \tag{10.31}$$

При $n > 0$ функция определена для $\mu > 0$. При $n = 0$ используем предел $n\ln(n/\mu) = 0$, и формула упрощается до

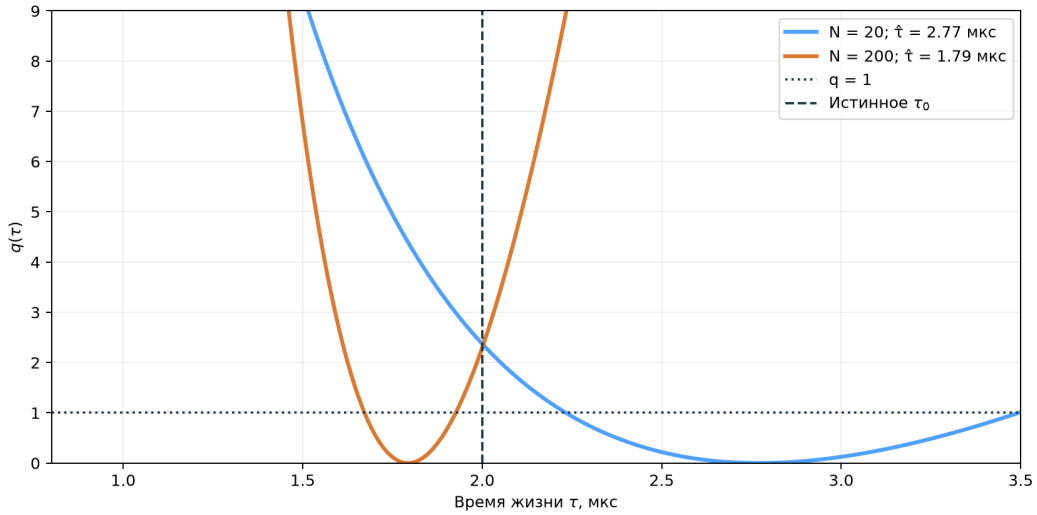


Рис. 10.1. *Отношение правдоподобий для первых 20 и всех 200 времён одной последовательности распадов при $\tau_0 = 2$ мкс. Горизонтальная линия задаёт $q = 1$, вертикальная — истинное время жизни.*

$$q(\mu) = 2\mu, \qquad \mu \geq 0. \tag{10.32}$$

Максимум L находится на границе $\hat{\mu} = 0$. Никакой параболы около внутреннего максимума здесь нет. Подстановка $n = 0$ в приближённую ошибку $\sqrt{n}$ дала бы ноль, хотя отсутствие событий не определяет μ с бесконечной точностью.

В таком случае сохраняем полную форму правдоподобия и выбираем процедуру построения интервала с нужным покрытием. К этим процедурам мы вернёмся в главах о статистическом выводе.

10.7. Несколько параметров

Пока у нас был один неизвестный параметр. В эксперименте вместе с ним обычно приходится определять фон, эффективность, энергетическую шкалу или разрешение детектора.

Разделим параметры по их роли в данном анализе. **Параметр интереса** θ — величина, которую мы хотим измерить. Остальные обозначим η и назовём **мешающими параметрами** (*nuisance parameters*). Например, при измерении сечения эффективность мешающая, а в калибровочном опыте из прошлой главы она сама была параметром интереса.

Сначала найдём совместный максимум:

$$(\hat{\theta}, \hat{\eta}) = \arg\max_{\theta,\eta} L(\theta, \eta). \tag{10.33}$$

Теперь хочется построить зависимость только от θ. Что делать с остальными параметрами? Если просто оставить их равными значениям в совместном максимуме, мы запретим им изменяться при изменении θ. При корреляциях это меняет вывод об ошибке.

10.7.1. Профильное правдоподобие

Зафиксируем пробное θ и заново найдём максимум по мешающим параметрам:

$$\hat{\hat{\eta}}(\theta) = \arg\max_{\eta} L(\theta, \eta). \tag{10.34}$$

Двойная шляпка обозначает условный максимум при выбранном θ. Повторяя эту операцию для каждого θ, получим **профильное правдоподобие**:

$$\begin{aligned} L_{\mathrm{p}}(\theta) &= L\bigl(\theta, \hat{\hat{\eta}}(\theta)\bigr), \\ q_{\mathrm{p}}(\theta) &= -2 \ln \frac{L_{\mathrm{p}}(\theta)}{L(\hat{\theta}, \hat{\eta})}. \end{aligned} \tag{10.35}$$

Это вполне конкретная численная процедура: каждому значению на горизонтальной оси соответствует своя подгонка мешающих параметров.

10.7.2. Двумерный гауссов пример

Перенесём начало координат в совместный максимум и измерим отклонения параметров в единицах их стандартных ошибок. Для гауссовой формы с корреляцией ρ, $|\rho| < 1$, имеем

$$q(\theta, \eta) = \frac{\theta^2 - 2\rho\theta\eta + \eta^2}{1 - \rho^2} = \theta^2 + \frac{(\eta - \rho\theta)^2}{1 - \rho^2}. \tag{10.36}$$

Вторая запись позволяет найти минимум по η. При фиксированном θ он достигается в точке

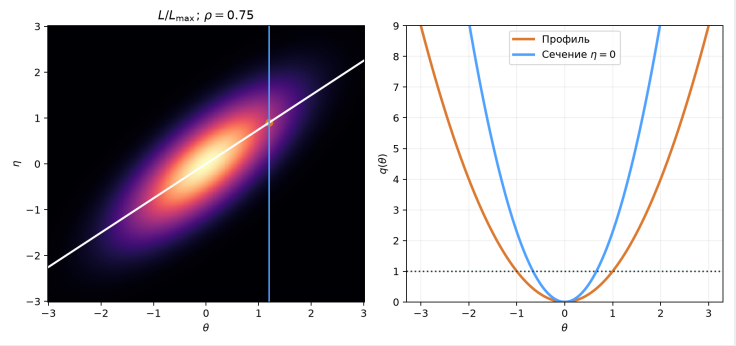
$$\hat{\hat{\eta}}(\theta) = \rho\theta, \qquad q_{\mathrm{p}}(\theta) = \theta^2. \tag{10.37}$$

А сечение при $\eta = 0$ равно

$$q(\theta, 0) = \frac{\theta^2}{1 - \rho^2}. \tag{10.38}$$

При $\rho = 0.75$ полуширина сечения по уровню $q = 1$ равна $\sqrt{1 - \rho^2} \simeq 0.66$, тогда как профиль даёт единицу. Фиксация мешающего параметра в его оценённом значении занизила ошибку примерно на треть. Если же η действительно известен точно из независимых соображений, фиксация оправданна: это уже другая постановка задачи.

В апплете белая линия на двумерной карте соединяет условные максимумы. Изменяйте ρ и следите за разницей между профилем и сечением. При $\rho = 0$ они совпадут.



Интерактив. Профиль и фиксированное сечение
Двумерное относительное правдоподобие при $\rho = 0.75$ и две одномерные кривые: профиль $q_{\mathrm{p}}(\theta)$ и сечение $q(\theta, 0)$. Белая линия показывает $\hat{\hat{\eta}}(\theta) = \rho\theta$.

10.8. Пример: реакторные нейтрино

Перейдём от гауссовой поверхности к физической модели. Детектор находится на расстоянии $L = 1600$ м от реактора. По энергетическому спектру зарегистрированных антинейтрино требуется оценить два осцилляционных параметра: амплитуду $A = \sin^2(2\theta)$ и разность квадратов масс Δm^2.

Возьмём двухфлейворную вакуумную вероятность выживания:

$$P_{ee}(E; A, \Delta m^2) = 1 - A \sin^2\!\left(1.267\, \Delta m^2 \frac{L}{E}\right). \tag{10.39}$$

В этой записи Δm^2 выражается в эВ^2, L — в метрах, E — в МэВ. Амплитуда задаёт глубину осцилляционного провала, а Δm^2 меняет его положение по энергии.

Это учебная модель. Будем считать расстояние, нормировку и спектр источника известными точно; фон и энергетическое разрешение пока отсутствуют. Измеряемая энергия здесь совпадает с энергией антинейтрино. В реальном детекторе наблюдается энергия продуктов взаимодействия, и для связи с E нужна модель отклика.

10.8.1. Ожидаемое число событий

Разобьём энергетический диапазон на бины. Если $\phi(E)$ описывает спектр источника, а $\sigma(E)$ — сечение регистрации, то среднее число событий в бине равно

$$\mu_i(A, \Delta m^2) = C \int_{E_i^-}^{E_i^+} \phi(E)\sigma(E) P_{ee}(E; A, \Delta m^2)\, dE. \tag{10.40}$$

В апплете интеграл приближён значением подынтегральной функции в центре бина, умноженным на его ширину:

$$\mu_i \simeq C\, \phi(E_i)\sigma(E_i) P_{ee}(E_i; A, \Delta m^2) \Delta E_i. \tag{10.41}$$

Одно и то же приближение используется при генерации и подгонке. Для анализа реальных данных точность интегрирования по бину проверяется отдельно.

Числа событий в непересекающихся бинах моделируем независимыми пуассоновскими величинами $n_i \sim \mathrm{Pois}(\mu_i)$. Тогда с точностью до константы

$$\ell(A, \Delta m^2) = \sum_i \left[n_i \ln \mu_i(A, \Delta m^2) - \mu_i(A, \Delta m^2) \right]. \tag{10.42}$$

Эту функцию и предстоит максимизировать.

10.8.2. Спектр источника и сечение

Используем параметризацию спектров Хубера и Мюллера [3, 4]:

$$\phi(E) = \sum_k f_k \exp\left[\sum_{p=0}^{5} a_{kp} (E/\text{МэВ})^p \right]. \tag{10.43}$$

Индекс k нумерует четыре делящихся изотопа, f_k — их доли делений. Общую размерную нормировку спектра включаем в C. Коэффициенты приведены в таблице 10.1. Доли выбраны для примера и во время расчёта не меняются. Продолжение параметризации ниже 2 МэВ также используется только как учебное приближение.

Таблица 10.1. Доли делений и коэффициенты спектров в учебной модели

Величина	^{235}U	^{238}U	^{239}Pu	^{241}Pu
f_k	0.58	0.07	0.30	0.05
a_0	4.367	0.4833	4.757	2.990
a_1	-4.577	0.1927	-5.392	-2.882
a_2	2.100	-0.1283	2.563	1.278
a_3	-0.5294	-0.006762	-0.6596	-0.3343
a_4	0.06186	0.002233	0.07820	0.03905
a_5	-0.002777	-0.0001536	-0.003536	-0.001754

Регистрация происходит через обратный бета-распад $\bar{\nu}_e + p \to e^+ + n$. В приближении без отдачи нуклонов энергетическая зависимость сечения имеет вид

$$\sigma(E) \propto E_e p_e, \qquad E_e = E - 1.293 \text{ МэВ}, \qquad p_e = \sqrt{E_e^2 - m_e^2}, \tag{10.44}$$

где $m_e = 0.511$ МэВ, $c = 1$. Ниже порога в модели полагаем сечение равным нулю. Поправки на отдачу здесь не учитываются. Диапазон расчёта составляет 1.8–8 МэВ, а C выбирается так, чтобы без осцилляций ожидаемое полное число событий равнялось заданному N_0.

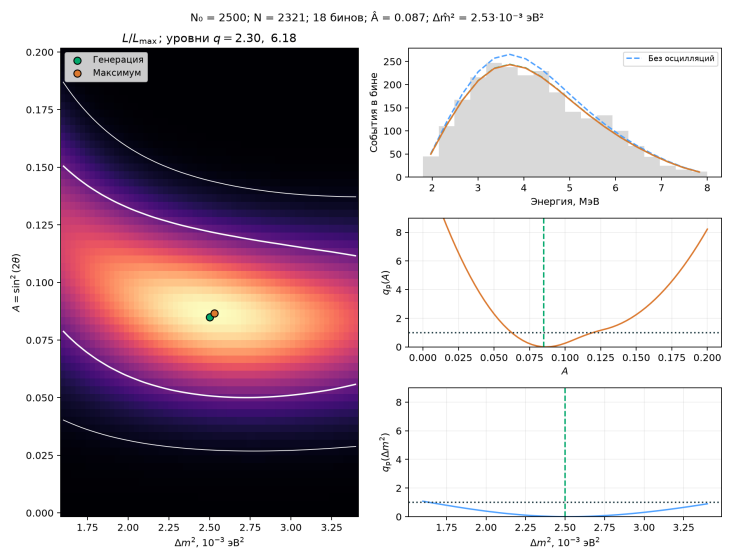
10.8.3. Генерация и подгонка

Сначала зададим истинные A_0 и Δm^2_0, вычислим средние числа событий и сгенерируем пуассоновские n_i. Полученные псевдоданные зафиксируем. Теперь переберём пробные пары $(A, \Delta m^2)$ и для каждой посчитаем логарифм правдоподобия 10.42.

В апплете используется сетка из 61×61 точки. Максимум и профили находятся именно на этой сетке:

$$\begin{aligned} q(A, \Delta m^2) &= -2[\ell(A, \Delta m^2) - \ell_{\max}], \\ q_{\mathrm{p}}(A) &= \min_{\Delta m^2} q(A, \Delta m^2), \\ q_{\mathrm{p}}(\Delta m^2) &= \min_{A} q(A, \Delta m^2). \end{aligned} \tag{10.45}$$

Здесь оба параметра представляют физический интерес. Но при построении одномерного профиля одного из них второй играет роль мешающего.



Интерактив. Подгонка реакторного спектра
Псевдоэксперимент при $N_0 = 2500$, $A_0 = 0.085$, $\Delta m_0^2 = 2.5 \cdot 10^{-3}$ эВ^2 и 18 энергетических бинах. Зелёным обозначены параметры генерации, оранжевым — максимум на сетке. Показаны двумерное правдоподобие, спектр событий и два одномерных профиля.

Начните с энергетического спектра. Насколько похожи кривые для истинных параметров и для найденного максимума? Затем посмотрите на двумерную карту. Разные пары параметров могут давать близкие спектры, поэтому область большого правдоподобия вытягивается. По одномерным профилям видно, насколько каждый параметр можно изменить, подстраивая второй.

На карте отмечены уровни $q = 2.30$ и $q = 6.18$. В регулярном гауссовом пределе они соответствуют совместным областям с покрытием примерно 68.3% и 95.4% для двух параметров. Уровень $q_\mathrm{p} = 1$ относится к одному профилируемому параметру. Разницу между одномерными и двумерными порогами мы уже видели при построении гауссовых эллипсов.

В этой модели есть место, где ссылка на гауссов предел особенно опасна. При $A = 0$ вероятность $P_{ee} = 1$ вообще не зависит от Δm^2: данные не могут определить частоту осцилляций, если их амплитуда равна нулю. Нарушается идентифицируемость параметра. Поэтому показанные уровни служат ориентирами; точность покрытия около $A = 0$ надо проверять псевдоэкспериментами.

Увеличьте N_0 и повторите генерацию. Статистический разброс уменьшается, детали спектра становятся лучше различимы. Если максимум оказался на краю сетки, проверьте сам диапазон поиска. Границы $A \le 0.20$ и $1.6 \le \Delta m^2/(10^{-3}\ \text{эВ}^2) \le 3.4$ выбраны для апплета. Они не являются полными физическими границами осцилляционных параметров.

10.9. Правдоподобие в реальном анализе

10.9.1. Вспомогательное измерение

Пусть неизвестный фон или калибровку η дополнительно измерили в независимом опыте. Результат a описывается моделью $a \sim \mathcal{N}(\eta, \sigma_a^2)$ с известной σ_a.

Тогда полное правдоподобие равно

$$L_{\text{полн}}(\theta, \eta) = L_{\text{осн}}(\theta, \eta) \exp\left[-\frac{(a-\eta)^2}{2\sigma_a^2}\right] \times \text{const.} \tag{10.46}$$

Гауссов множитель описывает распределение результата вспомогательного измерения при заданном η. Его происхождение то же, что у основной функции правдоподобия. При профилировании мы максимизируем произведение обоих множителей. В каждом условном максимуме учитываются и основные, и калибровочные данные.

10.9.2. Расширенное правдоподобие

Во многих экспериментах информация заключена одновременно в форме спектра и в полном числе событий. Пусть $f(x \mid \theta)$ — нормированная плотность признака одного события, а $\nu(\theta)$ — ожидаемое число зарегистрированных событий. При пуассоновском счёте и независимых событиях **расширенное правдоподобие** равно

$$L_{\text{расш}}(\theta) = \frac{e^{-\nu}\nu^N}{N!} \prod_{i=1}^{N} f(x_i \mid \theta). \tag{10.47}$$

Пуассоновский множитель использует число событий; произведение плотностей описывает их распределение по x.

Например, ожидается s сигнальных и b фоновых событий с нормированными плотностями f_s и f_b. Тогда $\nu = s + b$ и

$$f(x) = \frac{s f_s(x) + b f_b(x)}{s + b}.$$

После подстановки множители $(s+b)^N$ сокращаются:

$$L_{\text{расш}}(s, b) = \frac{e^{-(s+b)}}{N!} \prod_{i=1}^{N} [s f_s(x_i) + b f_b(x_i)]. \tag{10.48}$$

Можно добавить вспомогательное измерение фона и профилировать по b. Так получается модель, в которой число событий, форма спектра и измерение фона участвуют в одной подгонке.

10.9.3. Правдоподобие для гистограммы

Если сохранить только числа событий в бинах, расширенное правдоподобие переходит в произведение пуассоновских вероятностей:

$$L(\theta) = \prod_j \frac{e^{-\mu_j(\theta)} \mu_j(\theta)^{n_j}}{n_j!}. \tag{10.49}$$

Именно такую запись мы использовали для реакторного спектра. Независимость счётчиков соответствует пуассоновскому полному числу событий. Если полное число N фиксировано условием опыта, распределение чисел по бинам будет мультиномиальным: сумма $n_j = N$ связывает их между собой.

Гистограмма сокращает вычисления. Вместо миллионов событий можно работать с несколькими десятками или сотнями бинов. Цена слишком грубого разбиения — потеря информации о положении событий внутри бина. Например, узкий пик может целиком оказаться в одном широком бине: определить по нему положение пика будет трудно.

10.9.4. Замена параметров

Пусть нам удобнее подгонять $\phi = g(\theta)$, где g взаимно однозначна. Правдоподобие в новой переменной равно $L(g^{-1}(\phi))$, поэтому

$$\hat{\phi}_{\text{ММП}} = g(\hat{\theta}_{\text{ММП}}). \tag{10.50}$$

Это свойство называется **инвариантностью ММП**. Например, для скорости распада $\lambda = 1/\tau$ получаем $\hat{\lambda} = 1/\bar{t}$. Якобиан здесь не появляется: мы меняем аргумент функции правдоподобия. Преобразование плотности случайной величины — другая операция.

Несмещённость при нелинейной замене может потеряться, поскольку $\mathbb{E}[g(\hat{\theta})]$ в общем случае не равно $g(\mathbb{E}[\hat{\theta}])$. Границы интервала также надо преобразовывать самостоятельно. Для $\lambda = 1/\tau$ интервал $[\tau_-, \tau_+]$ переходит в $[1/\tau_+, 1/\tau_-]$.

10.10. Проверка численной подгонки

В простых примерах максимум мы нашли аналитически. В большой модели его ищет программа, и сообщение о сходимости само по себе не проверяет ни статистическую модель, ни глобальность максимума.

Начните с повторных запусков из разных начальных точек. Сравните значения логарифма правдоподобия и проверьте градиент во внутреннем максимуме. Затем постройте профили и двумерные сечения для коррелирующих параметров. Несколько типичных ситуаций собраны в таблице 10.2.

Таблица 10.2. Проверки численной подгонки

Что получилось	Что проверить
Разные результаты из разных начальных точек	Локальные максимумы и точность оптимизации
Максимум на краю диапазона поиска	Физическая ли это граница или только настройка программы
Почти плоский профиль	Чувствительность данных к параметру и возможное вырождение
Узкое сечение и широкий профиль	Корреляцию с мешающими параметрами
NaN или бесконечность	Допустимость параметров, вычисление логарифмов и нормировок

После этого повторите весь анализ на псевдоэкспериментах. Истинные параметры в них известны: можно измерить смещение оценки, её дисперсию и частоту попадания на границу. Сравните разброс оценок с ошибкой по кривизне. Если строите интервалы, посчитайте долю интервалов, накрывших истинное значение. Так проверяется покрытие всей принятой процедуры.

Псевдоэксперименты из той же модели проверяют работу метода внутри этой модели. Чтобы исследовать неверное описание фона или отклика детектора, нужно отдельно менять модель генерации. Подгонка к реальным данным требует также проверки согласия; ей будут посвящены следующие главы.

Итоги главы

- Средний квадрат ошибки складывается из дисперсии и квадрата смещения. Несмещённость, состоятельность и малый разброс — разные свойства оценки.
- Информация Фишера задаёт границу дисперсии несмещённых оценок при условиях регулярности. Обратная кривизна конкретного правдоподобия даёт локальную оценку ошибки; при малой выборке или на границе её надо проверять.
- При профилировании мешающие параметры заново подгоняются для каждого значения параметра интереса. Фиксация в совместном максимуме при корреляциях даёт другую, более узкую кривую.
- Псевдоэксперименты позволяют проверить смещение, разброс и покрытие интервалов. Асимптотические пороги отношения правдоподобий требуют отдельной проверки там, где нарушаются их условия применимости.

10.11. Задачи

Задача 1. Информация Фишера для гауссова среднего

Пусть

$$X_i \sim \mathrm{N}(\mu, \sigma^2),$$

где измерения независимы, а σ известна.

1. Вычислите информацию Фишера для одного наблюдения и всей выборки.
2. Получите границу Крамера–Рао для дисперсии несмещённой оценки μ.
3. Сравните границу с дисперсией выборочного среднего.

Задача 2. Смещённая оценка дисперсии

Для независимой гауссовой выборки размера $N \geq 2$ с неизвестными μ и σ^2:

1. Найдите оценки максимального правдоподобия обоих параметров.
2. Покажите, что
$$\mathbb{E}[\widehat{\sigma^2}_{\text{ММП}}] = \frac{N-1}{N}\sigma^2.$$
3. Постройте несмещённую оценку дисперсии.
4. Сравните средние квадраты ошибок двух оценок аналитически или методом Монте-Карло.
5. Обсудите, почему оценка максимального правдоподобия не обязана быть несмещённой при конечном N.

Задача 3. Профилирование мешающего параметра

Рассмотрите квадратичный логарифм функции правдоподобия

$$-2\ell(\theta,\eta) = \frac{\theta^2 - 2\rho\theta\eta + \eta^2}{1-\rho^2} + \text{const}, \qquad |\rho| < 1.$$

1. Найдите совместный максимум.
2. Для фиксированного θ найдите
$$\hat{\hat{\eta}}(\theta).$$
3. Постройте профильную функцию $q_{\mathrm{p}}(\theta)$.
4. Сравните её с сечением $q(\theta, \eta = 0)$.
5. Объясните, почему фиксация η в его оценке занижает неопределённость при $\rho \neq 0$. Как изменится постановка задачи, если η известен точно?

Задача 4. Расширенная функция правдоподобия для сигнала и фона

Наблюдаются события с признаком x. Ожидаемые числа сигнала и фона равны s и b, а нормированные плотности — $f_s(x)$ и $f_b(x)$.

1. Запишите расширенную функцию правдоподобия для выборки $x_1, \dots, x_N$.
2. Покажите, как в ней используются число событий и форма распределения.
3. Запишите функцию правдоподобия для бинированных данных.
4. Обсудите, какая информация теряется при грубом биннировании.
5. Независимое вспомогательное измерение фона дало a и описывается моделью $a \sim \mathcal{N}(b, \sigma_a^2)$ с известной σ_a. Добавьте его в полное правдоподобие и укажите, по какому параметру надо профилировать при оценивании s.

Задача 5. Проверка подгонки на псевдоэкспериментах

Выберите одну из моделей предыдущих задач.

1. Сгенерируйте не менее 5000 псевдоэкспериментов при известном истинном параметре.
2. В каждом псевдоэксперименте найдите оценку максимального правдоподобия.
3. Оцените смещение, дисперсию и средний квадрат ошибки.
4. Исследуйте зависимость результатов от размера выборки.
5. Если выбранная модель имеет физическую границу параметра, проверьте, как часто оценка попадает на неё.
6. Сравните фактический разброс оценки максимального правдоподобия с ошибкой, предсказанной кривизной логарифма функции правдоподобия.

Литература

[1] Cowan, G. *Statistical Data Analysis*. Oxford University Press. 1998.
[2] Wilks, S. S. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, **9**, 60–62. 1938. doi:10.1214/aoms/1177732360.
[3] Huber, P. Determination of Antineutrino Spectra from Nuclear Reactors. *Physical Review C*, **84**, 024617. 2011. doi:10.1103/PhysRevC.84.024617.
[4] Mueller, Th. A. и др. Improved Predictions of Reactor Antineutrino Spectra. *Physical Review C*, **83**, 054615. 2011. doi:10.1103/PhysRevC.83.054615.

Глава 11

Метод наименьших квадратов и линейная подгонка

Содержание

Аннотация

В этой главе мы применим метод максимального правдоподобия к измерениям с гауссовыми ошибками и получим метод наименьших квадратов. На примере прямой найдём оценки параметров, их ошибки и корреляцию. Затем разберём, с какой точностью известна сама подогнанная кривая и что можно узнать из остатков. Физическим примером послужит калибровка энергетической шкалы детектора.

11.1. Физическая задача: калибровка детектора

Детектор выдаёт амплитуду сигнала в каналах аналого-цифрового преобразователя, сокращённо АЦП. Для физического анализа нужна энергия. Чтобы связать одно с другим, измерим положение нескольких пиков с известными энергиями.

Рассмотрим учебный набор из трёх измерений в таблице 11.1. Энергии считаем известными точно. Ошибки положений пиков независимы, гауссовы и имеют известное стандартное отклонение 2 канала. Здесь речь об ошибке измеренного центра пика; ширина самого пика может быть значительно больше.

Таблица 11.1. Учебные данные для калибровки энергетической шкалы

Энергия x_i, МэВ	Положение пика y_i, канал	Ошибка σ_i, канал
1	103	2
2	200	2
3	303	2

Предположим, что в этом диапазоне средняя амплитуда линейно зависит от энергии:

$$f(x) = a + bx. \tag{11.1}$$

Параметр a задаёт смещение нуля в каналах, а b — коэффициент преобразования в каналах на МэВ. Подгонка, которую мы сейчас выведем, даёт

$$\hat{a} = (2.0 \pm 3.1)\ \text{канала}, \qquad \hat{b} = (100.0 \pm 1.4)\ \text{канала/МэВ}. \tag{11.2}$$

Двух ошибок для описания результата ещё недостаточно. Оценки $\hat{a}$ и $\hat{b}$ сильно антикоррелированы: коэффициент корреляции равен примерно -0.93. Если увеличить наклон, свободный член придётся уменьшить, чтобы прямая по-прежнему проходила вблизи точек.

Откуда взялись эти числа? Почему коррелируют параметры, хотя ошибки трёх измерений независимы? И как эту корреляцию учитывать, когда калибровка уже используется в анализе? Начнём с функции правдоподобия.

11.2. От правдоподобия к сумме квадратов

Пусть имеются N измерений $(x_i, y_i \pm \sigma_i)$ и модель $y = f(x;\theta)$. Во всей главе координаты x_i фиксированы и известны точно. Случайны результаты Y_i:

$$Y_i = f(x_i;\theta) + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, \sigma_i^2). \tag{11.3}$$

Ошибки ε_i независимы, а все $\sigma_i > 0$ известны и не зависят от параметров. Тогда правдоподобие имеет вид [1]

$$L(\theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{[y_i - f(x_i;\theta)]^2}{2\sigma_i^2}\right]. \tag{11.4}$$

Возьмём логарифм и умножим его на -2:

$$-2\ln L(\theta) = \sum_{i=1}^{N} \frac{[y_i - f(x_i;\theta)]^2}{\sigma_i^2} + 2\sum_{i=1}^{N} \ln(\sqrt{2\pi}\sigma_i). \tag{11.5}$$

Вторая сумма постоянна при изменении θ. Максимум правдоподобия поэтому совпадает с минимумом функции

$$\chi^2(\theta) = \sum_{i=1}^{N} \frac{[y_i - f(x_i;\theta)]^2}{\sigma_i^2}. \tag{11.6}$$

Это и есть **взвешенный метод наименьших квадратов**, или МНК.

Разность между измеренным значением и моделью называется **остатком**. Обозначим её через r_i. Вес точки равен обратной дисперсии:

$$r_i(\theta) = y_i - f(x_i;\theta), \qquad w_i = \frac{1}{\sigma_i^2}, \qquad \chi^2 = \sum_i w_i r_i^2. \tag{11.7}$$

Каждый вклад в χ^2 безразмерен. Отклонение на два канала для точки с ошибкой один канал даёт вклад 4. Для точки с ошибкой четыре канала тот же остаток даёт всего $1/4$. Так в подгонке учитывается разная точность измерений.

11.2.1. Что мы предположили об ошибках

Связь МНК с максимумом правдоподобия получена для конкретной модели 11.3. Если σ_i зависит от параметров, вторая сумма в формуле 11.5 тоже меняется, и отбрасывать её нельзя. При малом числе событий в бине следует использовать пуассоновское правдоподобие. Подстановка $\sigma_i = \sqrt{y_i}$ особенно неудачна для пустого бина: его вес тогда вообще не определён.

Сам метод наименьших квадратов можно применять и к негауссовым ошибкам. Ниже мы увидим, что несмещённость линейных оценок и формулы их ковариации следуют из первых двух моментов ошибок. Но совпадение с методом максимального правдоподобия и точные гауссовы интервалы в общем случае теряются.

Неопределённости по x и корреляции между измерениями требуют другого описания данных. В этой главе их пока нет. Физические ограничения на параметры тоже нужно учитывать отдельно: минимум может оказаться на границе, где обычные симметричные ошибки уже не описывают интервал.

11.3. Прямая по точкам

Вернёмся к модели $f(x) = a + bx$. Нужно минимизировать

$$\chi^2(a,b) = \sum_i w_i (y_i - a - bx_i)^2. \tag{11.8}$$

Слово «линейная» относится к зависимости от неизвестных параметров. Например, $a + bx + cx^2$ тоже линейная модель по a, b, c, хотя её график — парабола. Для начала достаточно двух параметров прямой.

11.3.1. Нормальные уравнения

Приравняем производные к нулю:

$$\begin{aligned} \frac{\partial \chi^2}{\partial a} &= -2 \sum_i w_i (y_i - a - b x_i) = 0, \\ \frac{\partial \chi^2}{\partial b} &= -2 \sum_i w_i x_i (y_i - a - b x_i) = 0. \end{aligned} \tag{11.9}$$

Введём пять сумм. Все они вычисляются по данным и известным весам:

$$\begin{aligned} S = \sum_i w_i, \qquad S_x = \sum_i w_i x_i, \qquad S_y = \sum_i w_i y_i, \\ S_{xx} = \sum_i w_i x_i^2, \quad S_{xy} = \sum_i w_i x_i y_i. \end{aligned} \tag{11.10}$$

Тогда два условия 11.9 образуют систему

$$\begin{pmatrix} S & S_x \\ S_x & S_{xx} \end{pmatrix} \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix}. \tag{11.11}$$

Это **нормальные уравнения** МНК. Шляпки у параметров появились потому, что мы уже ищем их оценки. Обозначим матрицу слева через A, вектор параметров через $\beta = (a, b)^{\mathsf{T}}$, а правую часть через $\mathbf{c}$. Решение записывается как $\hat{\beta} = A^{-1}\mathbf{c}$.

Для матрицы размера 2×2 обратную матрицу можно выписать сразу. С определителем $\Delta = S S_{xx} - S_x^2$ получаем

$$\hat{a} = \frac{S_{xx} S_y - S_x S_{xy}}{\Delta}, \qquad \hat{b} = \frac{S S_{xy} - S_x S_y}{\Delta}. \tag{11.12}$$

Численный поиск минимума здесь не требуется. При большом числе линейных параметров систему обычно решают средствами линейной алгебры, без явного вычисления обратной матрицы.

11.3.2. Можно ли определить наклон?

Введём взвешенный центр координат и меру их разброса:

$$\bar{x}_w = \frac{S_x}{S}, \qquad Q_x = \sum_i w_i (x_i - \bar{x}_w)^2 = S_{xx} - \frac{S_x^2}{S} = \frac{\Delta}{S}. \tag{11.13}$$

При положительных весах Q_x обращается в нуль только тогда, когда все x_i одинаковы. В этом случае мы измеряем значение прямой в одной точке. Через неё можно провести прямые с любым наклоном, меняя свободный член. Данные позволяют определить одну комбинацию $a + b x_1$, а двух отдельных параметров не определяют.

Если есть хотя бы два различных x_i, то $Q_x > 0$ и $\Delta > 0$. Матрица A положительно определена, поэтому найденная стационарная точка является единственным минимумом.

11.4. Ковариация оценок

В прошлой главе ошибка определялась по кривизне правдоподобия. Для прямой с известными гауссовыми ошибками этот способ даёт точный результат при любом числе точек, позволяющем определить оба параметра.

Положим $\delta a = a - \hat{a}$ и $\delta b = b - \hat{b}$. Раскрывая квадраты в формуле 11.8, получаем

$$\chi^2(a,b) - \chi^2_{\min} = S(\delta a)^2 + 2S_x \delta a \delta b + S_{xx}(\delta b)^2. \tag{11.14}$$

Линейные члены исчезли благодаря нормальным уравнениям. Членов третьей и более высокой степени нет: исходная функция уже была квадратичной. Следовательно, относительное правдоподобие равно

$$\frac{L(\beta)}{L(\hat\beta)} = \exp\!\left[-\frac{1}{2}(\beta-\hat\beta)^{\mathsf T} A (\beta-\hat\beta)\right]. \tag{11.15}$$

Матрица A играет роль обратной ковариации. Проверим это непосредственно по распределению оценок, чтобы не приписывать самому правдоподобию смысл распределения параметров.

Вывод

11.4.1. Дисперсии и ковариация из случайных ошибок

Вектор $\mathbf{c}$ из нормальных уравнений содержит две случайные суммы: $S_y = \sum_i w_i Y_i$ и $S_{xy} = \sum_i w_i x_i Y_i$. Подстановка $Y_i = a + bx_i + \varepsilon_i$ даёт

$$\begin{aligned} \mathbb{E}[\mathbf{c}] &= A\beta, \\ \operatorname{Cov}(\mathbf{c}) &= \begin{pmatrix} \sum_i w_i^2\sigma_i^2 & \sum_i w_i^2 x_i \sigma_i^2 \\ \sum_i w_i^2 x_i \sigma_i^2 & \sum_i w_i^2 x_i^2 \sigma_i^2 \end{pmatrix} = A. \end{aligned} \tag{11.16}$$

Использовано равенство $w_i\sigma_i^2 = 1$ и отсутствие ковариаций между ошибками разных точек. Поскольку $\hat\beta = A^{-1}\mathbf{c}$,

$$\mathbb{E}[\hat\beta] = \beta, \qquad V_\beta = \operatorname{Cov}(\hat\beta) = A^{-1} A (A^{-1})^{\mathsf T} = A^{-1}. \tag{11.17}$$

Обращение матрицы даёт

$$V_\beta = \frac{1}{\Delta}\begin{pmatrix} S_{xx} & -S_x \\ -S_x & S \end{pmatrix}. \tag{11.18}$$

Итак, оценки несмещённые. Их дисперсии и ковариация — три элемента матрицы 11.18:

$$\operatorname{Var}(\hat a) = \frac{S_{xx}}{\Delta}, \qquad \operatorname{Var}(\hat b) = \frac{S}{\Delta}, \qquad \operatorname{cov}(\hat a, \hat b) = -\frac{S_x}{\Delta}. \tag{11.19}$$

В этом выводе понадобились только средние и ковариации ошибок. Гауссова модель дополнительно гарантирует гауссово распределение вектора оценок: он является линейной функцией гауссовых данных.

Обратите внимание: в матрице V_β вообще нет значений y_i. При фиксированных x_i и известных σ_i случайные значения y_i меняют положение минимума, но его кривизна остаётся той же. Это свойство именно линейной модели с заданными ошибками.

11.5. Геометрия ошибок

Через $\bar x_w$ и Q_x формулы 11.19 переписываются в виде

$$\sigma_b^2 = \frac{1}{Q_x}, \qquad \sigma_a^2 = \frac{1}{S} + \frac{\bar x_w^2}{Q_x}, \qquad \operatorname{cov}(\hat a, \hat b) = -\frac{\bar x_w}{Q_x}. \tag{11.20}$$

Здесь σ_a и σ_b обозначают стандартные отклонения оценок. Чем шире разнесены точки по x при тех же ошибках, тем больше Q_x и тем точнее определяется наклон. Для

калибровки это означает: пики в узком энергетическом диапазоне плохо определяют коэффициент преобразования.

Со свободным членом ситуация другая. Параметр a — значение прямой при $x = 0$. Если калибровочные точки находятся далеко от нуля, приходится экстраполировать к нему прямую. Неопределённость наклона увеличивает неопределённость такого пересечения.

11.5.1. Через какую точку проходит прямая

Зафиксируем произвольный наклон b и найдём лучшее значение a. Первое нормальное уравнение даёт

$$\hat{a}(b) = \bar{y}_w - b\bar{x}_w, \qquad \bar{y}_w = \frac{S_y}{S}. \tag{11.21}$$

Двойная шляпка имеет тот же смысл, что и в прошлой главе: свободный член подобран при фиксированном наклоне. Подставляя $x = \bar{x}_w$, получаем $f(\bar{x}_w) = \bar{y}_w$. Все такие прямые проходят через взвешенный центр данных $(\bar{x}_w, \bar{y}_w)$.

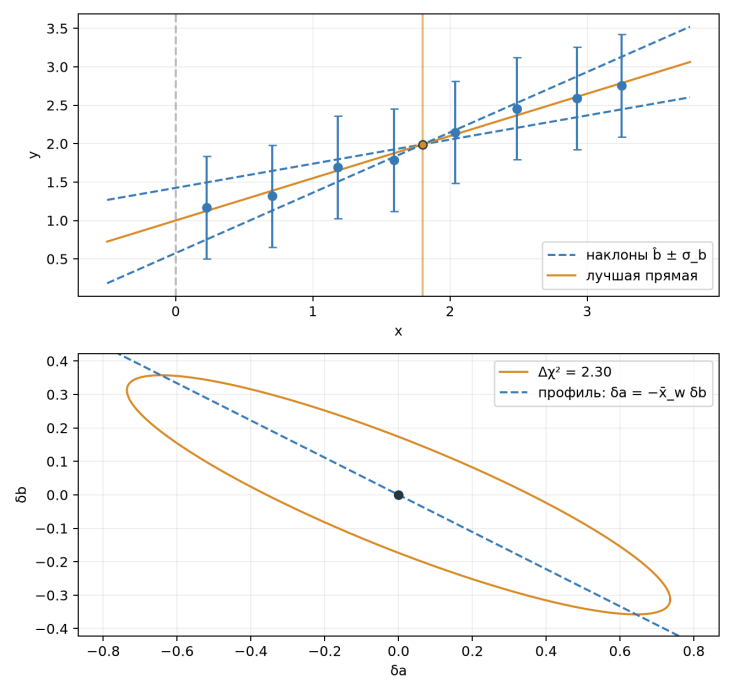
Теперь понятен знак ковариации. Если $\bar{x}_w > 0$, увеличение наклона компенсируется уменьшением свободного члена. При $\bar{x}_w < 0$ оба параметра меняются в одну сторону. Коэффициент корреляции равен

$$\rho_{ab} = \frac{\operatorname{cov}(\hat{a}, \hat{b})}{\sigma_a \sigma_b} = -\frac{S_x}{\sqrt{S S_{xx}}}. \tag{11.22}$$

11.5.2. Интерактив: центр точек и наклон

В апплете можно менять взвешенный центр $\bar{x}_w$ и среднеквадратичный разброс координат $s_x = \sqrt{Q_x/S}$. Суммарный вес $S = 18$ остаётся фиксированным. Координаты здесь безразмерные. Учебные точки построены так, чтобы их лучшая прямая была $y = 1 + 0.55x$.

Сплошная оранжевая линия показывает эту прямую. Две синие пунктирные линии соответствуют наклонам $\hat{b} \pm \sigma_b$ с заново подобранным свободным членом. На втором графике изображён контур $\Delta\chi^2 = 2.30$ и направление профилирования $\delta a = -\bar{x}_w \delta b$.



Интерактив. Геометрия корреляции параметров
При $\bar{x}_w = 1.8$ и $s_x = 1$ корреляция оценок равна -0.87. Перенос центра к нулю устраняет наклон эллипса. Увеличение разброса точек уменьшает ошибку наклона прямой.

11.5.3. Центрированная параметризация

Можно сразу измерять x от центра данных:

$$x' = x - \bar{x}_w, \qquad f(x) = a' + bx', \qquad a' = a + b\bar{x}_w. \tag{11.23}$$

Теперь $\sum_i w_i x'_i = 0$, поэтому

$$\hat{a}' = \bar{y}_w, \qquad \mathrm{Var}(\hat{a}') = \frac{1}{S}, \qquad \mathrm{cov}(\hat{a}', \hat{b}) = 0. \tag{11.24}$$

Наклон и его ошибка сохранились. Параметр a' теперь описывает значение прямой в центре данных. Сама калибровка не изменилась, но её параметры стали некоррелированными. В гауссовой модели оценки $\hat{a}'$ и $\hat{b}$ также независимы.

11.5.4. Нормальные координаты и профиль

В новых координатах квадратичная форма диагональна:

$$\Delta\chi^2 = S(\delta a')^2 + Q_x(\delta b)^2 = z_a^2 + z_b^2, \qquad z_a = \sqrt{S}\,\delta a', \quad z_b = \sqrt{Q_x}\,\delta b. \tag{11.25}$$

Мы получили те же нормальные координаты, которые вводили для двумерного гаусса. Эллипсы превратились в окружности. Здесь преобразование точно во всей плоскости параметров.

При профилировании по a' первое слагаемое обращается в нуль. Остаётся

$$\chi^2_{\mathrm{p}}(b) - \chi^2_{\min} = Q_x(b - \hat{b})^2 = \frac{(b - \hat{b})^2}{\sigma_b^2}. \tag{11.26}$$

Поэтому пределы $\hat{b} \pm \sigma_b$ соответствуют росту профильного χ^2 на единицу. Это одномерный интервал с покрытием около 68.3% в нашей гауссовой модели. Для совместной области двух параметров с тем же покрытием нужен порог $\Delta\chi^2 = 2.30$. Именно он показан в апплете.

11.6. Возвращаемся к калибровке

Для данных из таблицы 11.1 все веса равны $1/4$ в обратных квадратах канала. Получаем

$$\bar{x}_w = 2 \text{ МэВ}, \qquad \bar{y}_w = 202 \text{ канала}, \qquad Q_x = 0.5 \text{ МэВ}^2/\text{канал}^2. \tag{11.27}$$

Из формул 11.12 и 11.20 следуют $\hat{b} = 100$ каналов/МэВ и $\hat{a} = 2$ канала, $\sigma_b = \sqrt{2}$ канала/МэВ и $\sigma_a = \sqrt{28/3}$ канала. Ковариация равна -4 канал^2/МэВ. Это и даёт результат 11.2 с корреляцией $-\sqrt{6/7} \simeq -0.93$.

В центрированной записи та же прямая выглядит проще:

$$\widehat{f}(x) = 202 \text{ канала} + \left(100 \frac{\text{каналов}}{\text{МэВ}}\right)(x - 2 \text{ МэВ}). \tag{11.28}$$

Значение в центре равно 202 ± 1.15 канала и не коррелирует с наклоном. Ошибка 1.15 заметно меньше ошибки свободного члена 3.1: центр данных находится внутри диапазона измерений, а нулевая энергия — за его пределами.

11.6.1. С какой точностью известна прямая?

В произвольной фиксированной точке x оценка среднего сигнала равна $\widehat{f}(x) = \hat{a} + \hat{b}x$. Закон распространения ошибок даёт

$$\begin{aligned} \operatorname{Var}[\widehat{f}(x)] &= \sigma_a^2 + 2x \operatorname{cov}(\hat{a}, \hat{b}) + x^2 \sigma_b^2 \\ &= \frac{1}{S} + \frac{(x - \bar{x}_w)^2}{Q_x}. \end{aligned} \tag{11.29}$$

Это точное равенство: $\widehat{f}$ линейна по оценкам. Минимальная ошибка достигается в центре данных. По мере удаления от центра полоса вокруг подогнанной прямой расширяется.

Если забыть ковариацию, уже при $x = 2$ МэВ получится дисперсия $28/3 + 8 = 52/3$ канал^2 вместо $4/3$ канал^2. Ошибка вырастет с 1.15 до 4.16 канала. Сильная корреляция здесь имеет вполне измеримое последствие.

11.6.2. Средний сигнал и новое измерение

Теперь проведём ещё одно независимое измерение при фиксированной энергии x_*, с гауссовой ошибкой известной ширины σ_*:

$$Y_* = a + bx_* + \varepsilon_*, \qquad \operatorname{Var}[Y_* - \widehat{f}(x_*)] = \sigma_*^2 + \frac{1}{S} + \frac{(x_* - \bar{x}_w)^2}{Q_x}. \tag{11.30}$$

Первое слагаемое описывает флуктуацию нового измерения. Два остальных связаны с неопределённостью калибровочной прямой. Поэтому интервал для нового измерения шире интервала для среднего сигнала.

В центре нашей калибровки средний сигнал известен с ошибкой 1.15 канала. Если новое измерение имеет $\sigma_* = 2$ канала, стандартное отклонение разности $Y_* - \widehat{f}(x_*)$ равно $\sqrt{4 + 4/3} = 2.31$ канала. Обе полосы показаны на рисунке 11.1.

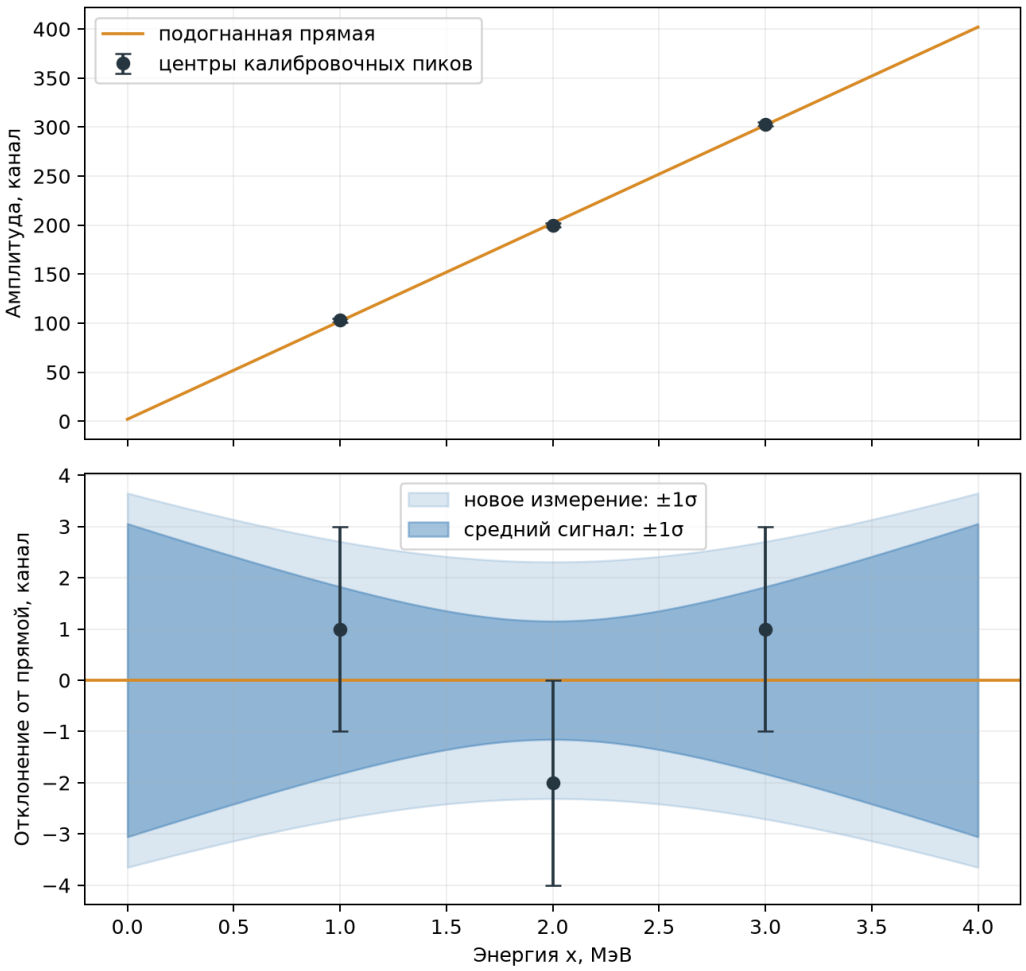


Рис. 11.1. *Учебная калибровка из таблицы: точки с ошибками и подогнанная прямая. На нижнем графике из каждой величины вычтена подогнанная прямая, чтобы различить полосы. Узкая полоса соответствует одной стандартной ошибке среднего сигнала; широкая учитывает также ошибку нового независимого измерения $\sigma_* = 2$ канала.*

Границы обеих полос имеют точечное покрытие около 68.3% при каждом заранее фиксированном x. Утверждение о всей кривой сразу требует одновременной доверительной полосы с другим порогом. Кроме того, за пределами калибровочного диапазона эти формулы учитывают лишь ошибки параметров прямой. Возможную нелинейность реального отклика они не описывают.

11.7. Что остаётся после подгонки

Остатки при найденных параметрах равны

$$r_i = y_i - \hat{a} - \hat{b}x_i, \qquad u_i = \frac{r_i}{\sigma_i}. \tag{11.31}$$

Величины u_i позволяют сравнивать отклонения точек с разными ошибками. Однако считать их независимой выборкой из $\mathcal{N}(0,1)$ нельзя. Прямая уже подстроилась под те же точки. Из нормальных уравнений следуют два ограничения:

$$\sum_i w_i r_i = 0, \qquad \sum_i w_i x_i r_i = 0. \tag{11.32}$$

Изменение одной точки меняет прямую и тем самым остатки остальных точек.

Найдём эту связь явно.

11.7.1. Матрица влияния

Разделим данные и модель в каждой строке на σ_i:

$$z_i = \frac{y_i}{\sigma_i}, \qquad B_{i1} = \frac{1}{\sigma_i}, \qquad B_{i2} = \frac{x_i}{\sigma_i}. \tag{11.33}$$

Матрица B имеет N строк и два столбца. Ошибки вектора $\mathbf{z}$ независимы и имеют единичную дисперсию. Нормальные уравнения теперь записываются как $B^{\mathsf{T}} B\hat{\beta} = B^{\mathsf{T}}\mathbf{z}$. Значения подогнанной модели в этих координатах равны

$$\hat{\mathbf{z}} = H\mathbf{z}, \qquad H = B(B^{\mathsf{T}}B)^{-1}B^{\mathsf{T}}. \tag{11.34}$$

Матрица H называется **матрицей влияния**. Она симметрична и удовлетворяет $H^2 = H$: если ещё раз подогнать уже лежащие на прямой значения, они не изменятся. Вектор нормированных остатков равен $\mathbf{u} = (I - H)\mathbf{z}$, где I — единичная матрица. Поэтому

$$\operatorname{Cov}(\mathbf{u}) = (I - H)(I - H)^{\mathsf{T}} = I - H. \tag{11.35}$$

В частности,

$$\operatorname{Var}(u_i) = 1 - h_{ii}, \qquad h_{ii} = w_i\left[\frac{1}{S} + \frac{(x_i - \bar{x}_w)^2}{Q_x}\right]. \tag{11.36}$$

Точка с большим h_{ii} сильнее подтягивает к себе прямую, поэтому её остаток имеет меньшую дисперсию. При прочих равных большое влияние имеют точки далеко от центра x. Малый остаток такой точки сам по себе ещё не означает, что с измерением всё в порядке.

Для известной дисперсии ошибки и $h_{ii} < 1$ величина $u_i/\sqrt{1 - h_{ii}}$ имеет единичную дисперсию, а в нашей гауссовой модели и стандартное нормальное распределение. Эта нормировка не устраняет корреляций между разными остатками.

11.7.2. Почему остаётся $N - 2$ степеней свободы

Матрица H проектирует данные на двумерное пространство моделей, заданное двумя столбцами B. У неё два собственных значения 1, остальные равны 0, и $\operatorname{tr} H = 2$. Для остатков остаётся $N - 2$ независимых направлений. При верной линейной модели и известных гауссовых ошибках

$$\chi^2_{\min} = \mathbf{u}^{\mathsf{T}}\mathbf{u} \sim \chi^2_{N-2}, \qquad \mathbb{E}[\chi^2_{\min}] = N - 2. \tag{11.37}$$

Распределение справа обозначено тем же символом χ^2, что и функция, которую мы минимизировали. Здесь $N > 2$. Для нашей калибровки остатки равны $1, -2, 1$ канала, поэтому $\chi^2_{\min} = 1.5$ при одной степени свободы. Из одного такого числа нельзя заключить, что модель детектора верна. Формальную проверку согласия разберём в главе 13.

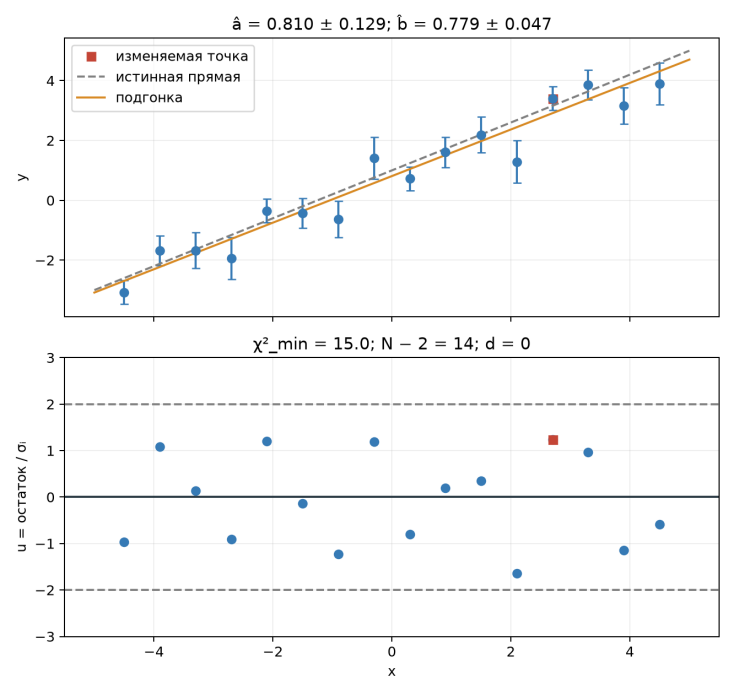
Если ошибки всех точек имеют одинаковую, но неизвестную дисперсию σ^2, её можно оценить по остаткам:

$$s^2 = \frac{1}{N-2}\sum_i r_i^2, \qquad \mathbb{E}[s^2] = \sigma^2. \tag{11.38}$$

Для интервалов тогда требуется учитывать случайность s, что приводит к распределению Стьюдента. В нашей основной задаче все σ_i заданы заранее. Перемасштабировать их до $\chi^2_{\min}/(N-2) = 1$ оснований нет.

11.8. Остатки и выбросы

Сгенерируем 16 точек около прямой $y = 1 + 0.8x$ с независимыми гауссовыми ошибками. Обе координаты здесь безразмерные. В апплете можно изменить масштаб ошибок и добавить сдвиг d к одной отмеченной точке. При изменении сдвига остальные точки сохраняются. Кнопка «Новые данные» меняет всю случайную выборку.



Интерактив. Прямая, остатки и выброс
Подгонка 16 псевдоизмерений при масштабе ошибок 0.55 и без дополнительного сдвига. Пунктиром показана истинная прямая, сплошной линией — результат подгонки. Нижний график содержит остатки, делённые на ошибки точек.

На рисунке 11.2 к той же выборке добавлен сдвиг $d = 4$ в одной точке. Её заявленная ошибка не изменилась. Подгонка учитывает эту точку с прежним весом, и прямая смещается.

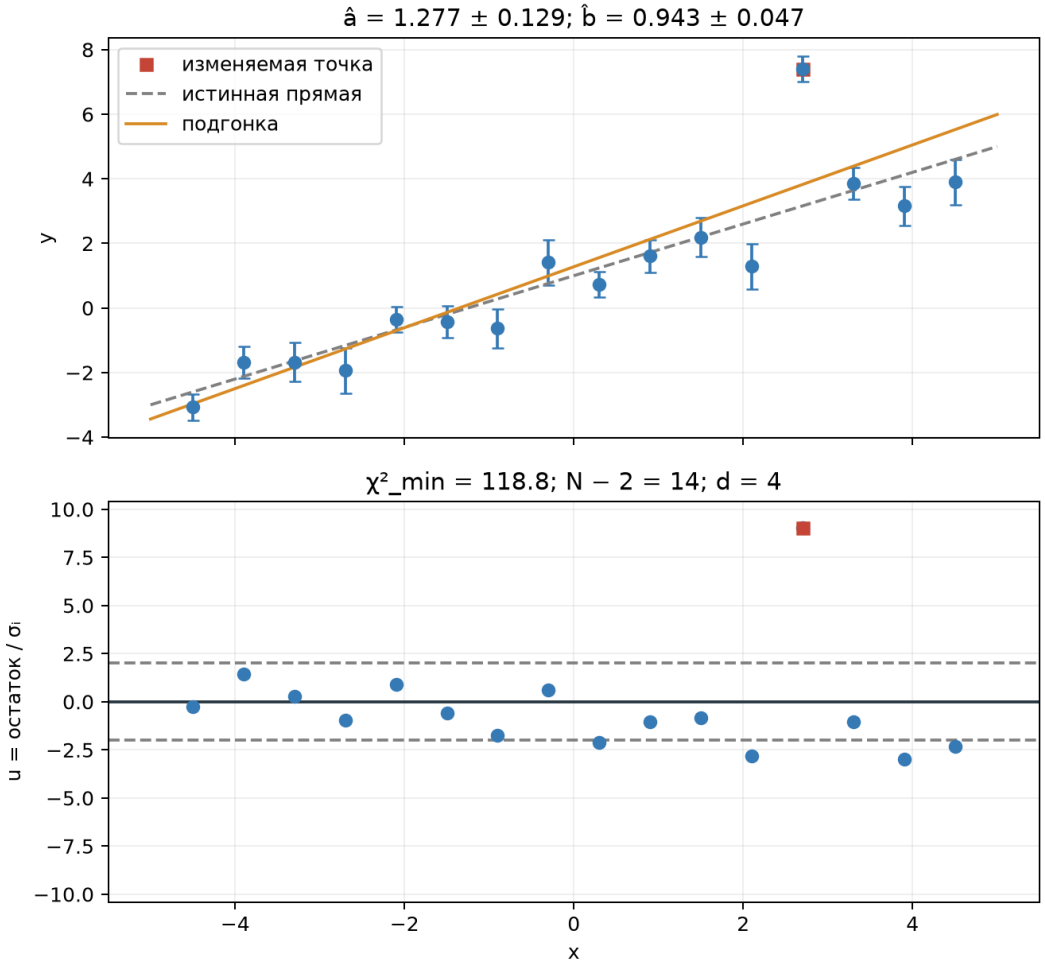


Рис. 11.2. *Те же псевдоизмерения после сдвига отмеченной точки на $d = 4$. Пунктирная прямая задаёт исходную модель, сплошная получена из подгонки. Сдвиг одной точки изменил и её собственный остаток, и остатки остальных измерений.*

Этот эффект можно вычислить без повторной минимизации. Если заменить y_j на $y_j + d$ при неизменных координатах и весах, то

$$\delta\hat{b} = \frac{w_j(x_j - \bar{x}_w)}{Q_x}\, d, \qquad \delta\hat{a} = \frac{w_j}{S}\, d - \bar{x}_w \delta\hat{b}. \tag{11.39}$$

Смещение наклона растёт линейно с d и с расстоянием точки от центра. Обычный МНК не ограничивает влияние сколь угодно большого отклонения.

Что делать, если такая точка встретилась в эксперименте? Нужно проверить причину: сбой электроники, неверную ошибку измерения, примесь другого процесса, насыщение детектора или недостаточность самой модели. Удаление точки требует физического основания и явно описанного правила отбора. Если большие отклонения составляют часть распределения ошибок, это распределение следует включить в правдоподобие.

На графике остатков полезно смотреть также на согласованное поведение многих точек. Изгиб может указывать на нелинейность отклика. Рост разброса с x может означать, что зависимость ошибок от энергии описана неверно. Такой график помогает поставить вопрос к модели; сам по себе он ещё не выбирает причину отклонения.

Итоги главы

- Для независимых гауссовых измерений с известными, не зависящими от параметров дисперсиями МНК совпадает с методом максимального правдоподобия.
- Подгонка прямой сводится к двум нормальным уравнениям. Ковариация оценок определяется координатами точек и их ошибками.
- Центрирование координаты устраняет ковариацию наклона и значения прямой в центре данных. При вычислении ошибки кривой в исходных параметрах ковариационный член необходимо сохранять.
- После подгонки остатки связаны между собой и имеют уменьшенные дисперсии. Их структура и влияние отдельных точек позволяют проверять описание измерений.

11.9. Задачи

Задача 1. Среднее как результат МНК

Имеются независимые измерения одной величины:

$$Y_i \sim \mathrm{N}(\mu, \sigma_i^2).$$

Все $\sigma_i > 0$ известны и не зависят от μ.

1. Запишите функцию $\chi^2(\mu)$.
2. Найдите её минимум.
3. Покажите, что полученная оценка является взвешенным средним.
4. Найдите дисперсию этой оценки.
5. Рассмотрите частный случай одинаковых σ_i.

Задача 2. Аналитическая подгонка прямой

Для данных (x_i, y_i, σ_i) рассматривается модель

$$y = a + bx.$$

Координаты x_i известны точно, ошибки по y независимы и гауссовы, их стандартные отклонения $\sigma_i > 0$ известны. Среди x_i есть хотя бы два различных значения.

1. Выведите нормальные уравнения.
2. Получите аналитические формулы для $\hat{a}$ и $\hat{b}$.
3. Выведите ковариационную матрицу параметров.
4. Покажите, что центрирование x_i около взвешенного среднего устраняет ковариацию параметров.
5. Проверьте формулы численно на искусственной выборке.

Задача 3. Ошибка подогнанной кривой

Три независимых гауссовых измерения при точно известных энергиях $x = (1, 2, 3)$ МэВ дали положения пиков $y = (103, 200, 303)$ канала. Ошибка каждого положения равна 2 каналам. Для модели $f(x) = a + bx$ найдите:

1. дисперсию оценки среднего $\hat{y}(x) = \hat{a} + \hat{b}x$;
2. точку, в которой эта дисперсия минимальна;
3. дисперсию разности $Y_* - \hat{y}(x_*)$ для нового независимого измерения при фиксированном x_* с известной ошибкой σ_*;
4. различие между полосой среднего сигнала и полосой для нового измерения.

Постройте обе полосы для этих данных при $\sigma_* = 2$ канала. Объясните, почему точечное покрытие не означает покрытия всей кривой сразу.

Задача 4. Остатки и диагональные элементы матрицы влияния

Пусть $N > 2$, $y_i = a + bx_i + \varepsilon_i$, а ε_i независимы, имеют нулевые средние и одну известную дисперсию σ^2. Координаты x_i фиксированы и не все одинаковы. В этом случае $\hat{\mathbf{y}} = H\mathbf{y}$, где $H = X(X^{\mathsf{T}}X)^{-1}X^{\mathsf{T}}$, $X_{i1} = 1$, $X_{i2} = x_i$. Обозначьте остатки через $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}$, а единичную матрицу через I.

1. покажите, что $H^2 = H$;
2. покажите, что $\operatorname{tr} H = 2$;
3. выведите ковариацию остатков

$$\operatorname{Cov}(\mathbf{r}) = \sigma^2(I - H);$$

4. найдите диагональные элементы h_{ii}, характеризующие влияние точек при подгонке прямой;
5. добавьте одну далёкую по x точку и исследуйте её влияние.

Задача 5. Влияние одной точки

Сгенерируйте 16 независимых гауссовых измерений около прямой $y = 1 + 0.8x$ при $x_i = -4.5 + 9(i-1)/15$, $i = 1, \dots, 16$, и известных одинаковых ошибках $\sigma_i = 0.55$. Зафиксируйте начальное состояние генератора.

1. Выполните подгонку обычным методом наименьших квадратов.
2. Добавьте к одной точке сдвиг d, сохранив её заявленную ошибку. Выведите изменения $\hat{a}$ и $\hat{b}$ из нормальных уравнений.
3. Меняйте d от -4 до 6 и сравните формулы с повторной подгонкой.
4. Повторите опыт для точки вблизи центра и для крайней точки. Сравните влияние на наклон и остатки.
5. Проверьте, меняется ли вычисленная ковариация параметров при этих сдвигах. Объясните, почему неизменная ошибка параметра не гарантирует отсутствия смещения из-за испорченного измерения.

Литература

[1] Cowan, G. *Statistical Data Analysis.* Oxford University Press. 1998.

Глава 12

χ^2-интервалы и систематические неопределённости

Содержание

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:

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

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

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

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:

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

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

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

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:

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

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:

$$\lambda = X\theta, \qquad X \in \mathbb{R}^{N\times p}. \tag{12.6}$$

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

$$X^{\mathsf{T}} V_y^{-1} X\hat\theta = X^{\mathsf{T}} V_y^{-1}\mathbf{y}. \tag{12.7}$$

Если столбцы X линейно независимы, решение единственно:

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

Проверка ковариации повторяет вывод для прямой. Оценка имеет вид $\hat\theta = B\mathbf{Y}$, где $BX = I$. Поэтому $\mathbb{E}[\hat\theta] = \theta$ и $\operatorname{Cov}(\hat\theta) = BV_yB^{\mathsf{T}} = V_\theta$. 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:

$$\begin{aligned} \chi^2(\theta) &\simeq \chi^2_{\min} + (\theta-\hat\theta)^{\mathsf{T}} K (\theta-\hat\theta), \\ K_{ij} &= \frac{1}{2}\left.\frac{\partial^2\chi^2}{\partial\theta_i\partial\theta_j}\right|_{\hat\theta}, \qquad V_\theta \simeq K^{-1}. \end{aligned} \tag{12.9}$$

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:

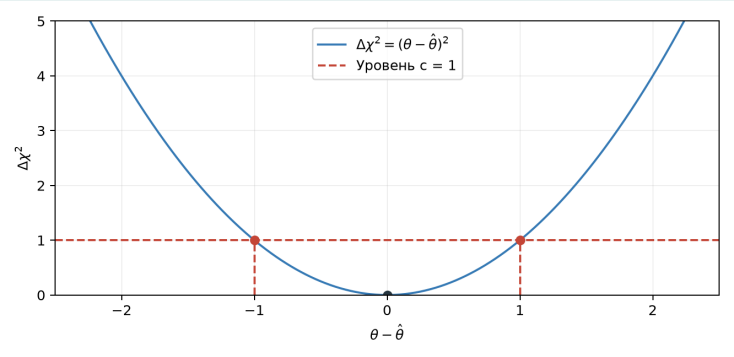
$$\Delta\chi^2(\theta) = \chi^2(\theta) - \chi^2_{\min} \simeq \frac{(\theta - \hat{\theta})^2}{\sigma_\theta^2}. \tag{12.10}$$

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

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

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



Интерактив. Интервал по пересечению с уровнем χ². При $\sigma_\theta = 1$ 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

$$(\theta - \hat{\theta})^{\mathsf{T}} V_\theta^{-1} (\theta - \hat{\theta}) \le c \tag{12.12}$$

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.

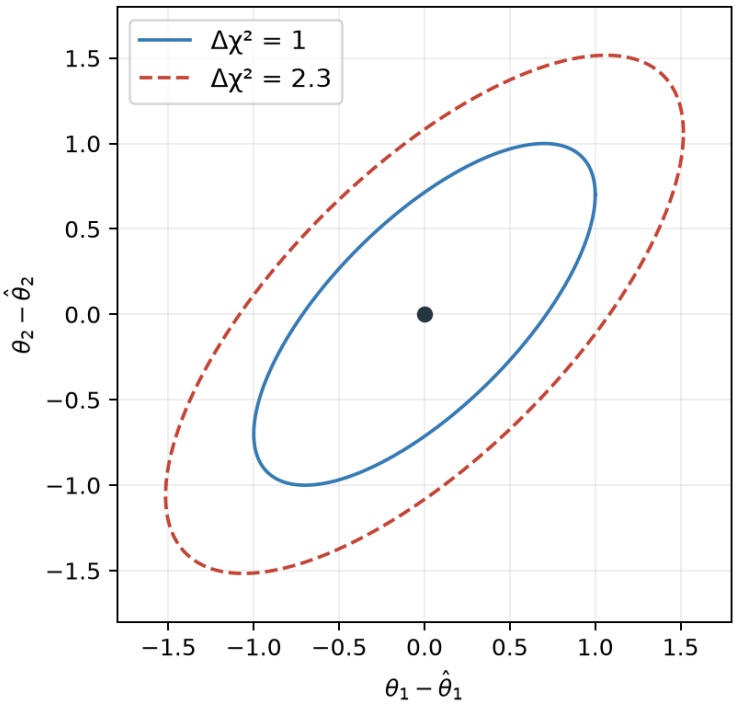


Рис. 12.1. *Контуры квадратичной формы для двух параметров с единичными стандартными ошибками и корреляцией $\rho = 0.7$. Сплошной контур соответствует $c = 1$, пунктирный — $c = 2.30$. Отложены отклонения параметров от минимума.*

Порог для совместной области зависит от числа параметров. Для двух гауссовых оценок область с $c = 1$ покрывает истинную пару всего примерно в 39.3% опытов. Для покрытия 68.3% требуется $c \simeq 2.30$. 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 $y_i = 0.6 \sin(1.2 t_i)$. Это специально заданные точки без добавленной случайной флуктуации. Приписав им независимые гауссовы ошибки $s = 0.15$, получим функцию

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

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.

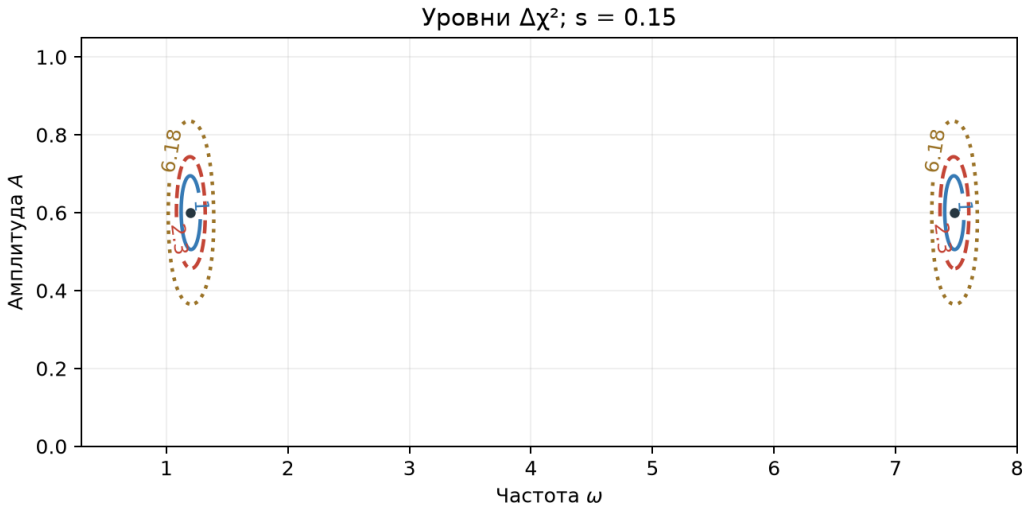


Рис. 12.2. *Контуры $\Delta\chi^2 = 1, 2.30$ и 6.18 для четырёх точек осцилляционной модели. Минимумы при $A = 0.6$, $\omega = 1.2$ и $\omega = 1.2 + 2\pi$ дают одинаковые предсказания. Локальный эллипс около одного минимума не описывает вторую область.*

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:

$$\lambda = \lambda(\theta, \eta). \tag{12.14}$$

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

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

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

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

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:

$$\chi^2_{\mathrm{p}}(\theta) = \min_{\eta} \chi^2(\theta,\eta) = \chi^2\bigl(\theta, \hat{\hat{\eta}}(\theta)\bigr). \tag{12.17}$$

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:

$$K = \begin{pmatrix} K_{\theta\theta} & K_{\theta\eta} \\ K_{\eta\theta} & K_{\eta\eta} \end{pmatrix}, \qquad K_{\mathrm{p}} = K_{\theta\theta} - K_{\theta\eta} K_{\eta\eta}^{-1} K_{\eta\theta}. \tag{12.18}$$

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 $\mathbf{z}$:

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

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

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

Производная по δ даёт систему

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

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

$$\chi^2_{\mathrm{p}} = \mathbf{r}^{\mathsf{T}} \left(W - W H M^{-1} H^{\mathsf{T}} W \right) \mathbf{r}. \tag{12.22}$$

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

$$\chi^2_{\mathrm{p}}(\theta) = \mathbf{r}^{\mathsf{T}} V^{-1} \mathbf{r}, \qquad V = V_y + H V_\eta H^{\mathsf{T}}. \tag{12.23}$$

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:

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

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 $1 + \alpha \le 0$.

Полная функция для подгонки равна

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

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

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

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

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

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

$$\hat\alpha(\mu) = \frac{f_N^2 \mu (m - \mu)}{v + f_N^2 \mu^2}, \qquad \chi^2_{\mathrm{p}}(\mu) = q_0 + \frac{(m-\mu)^2}{v + f_N^2\mu^2}. \tag{12.28}$$

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

$$\sigma_\mu^2 \simeq v + f_N^2 m^2. \tag{12.29}$$

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

$$m = 8.23486 \text{ пб}, \qquad v = 0.013575 \text{ пб}^2, \qquad f_N^2 m^2 = 0.678130 \text{ пб}^2. \tag{12.30}$$

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

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

Верхний левый элемент имеет размерность пб^2, внедиагональный — пб, нижний правый безразмерный. Отрицательная ковариация соответствует модели: увеличение масштаба компенсируется уменьшением сечения.

Для границ по $\Delta\chi^2_{\mathrm{p}} = 1$ можно использовать весь профиль 12.28. При $f_N < 1$ решение имеет вид

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

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

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

Весь нормировочный вклад пропорционален матрице из единиц. Подстановка этой матрицы в $\mathbf{r}^{\mathsf{T}} V(\mu)^{-1} \mathbf{r}$ 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 $y_i y_j$:

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

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

$$\hat{\mu} = \mathbf{w}^{\mathsf{T}} \mathbf{y}, \qquad \mathbf{w} = \frac{V^{-1}\mathbf{1}}{\mathbf{1}^{\mathsf{T}} V^{-1} \mathbf{1}}, \qquad \sigma_\mu^2 = \frac{1}{\mathbf{1}^{\mathsf{T}} V^{-1} \mathbf{1}}. \tag{12.35}$$

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:

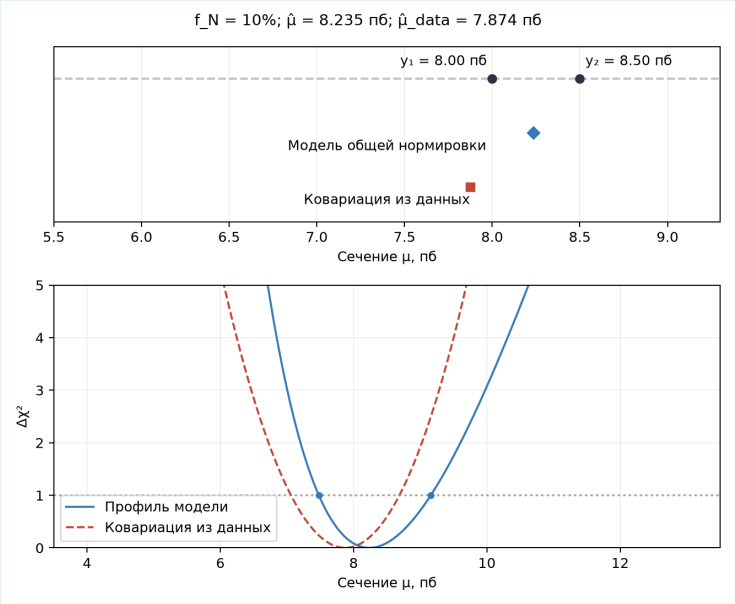
$$\hat{\mu}_{\text{data}} = \frac{m}{1 + f_N^2 q_0}, \qquad q_0 = \frac{(y_1 - y_2)^2}{s_1^2 + s_2^2}. \tag{12.36}$$

Чем сильнее расходятся точки, тем больше знаменатель. Для положительного m такая конструкция уменьшает центральное значение. В нашем примере $q_0 \simeq 4.59$, 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 m. 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:

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

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 $\widehat{E} = (y_* - \hat{a})/\hat{b}$. Если измерение y_* независимо от калибровки и имеет дисперсию s_*^2, то в линейном приближении

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

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 $V_y + HV_\eta H^\mathsf{T}$.
- 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

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

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 $\rho \to 0$, $\rho \to 1$ и $\rho \to -1/2$ 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 $V = s^2 I + \tau^2 \mathbf{1}\mathbf{1}^{\mathsf{T}}$, где I — единичная матрица, $\mathbf{1}$ — вектор из единиц.
4. Что происходит с ошибкой $\hat{\mu}$ при $N \to \infty$? 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 $\widehat{A}$.
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 $\Delta\chi^2_{\mathrm{p}} = 1$. 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, $\mathrm{Var}(\hat{a}) = 28/3$ канал^2, $\mathrm{Var}(\hat{b}) = 2$ канал^2/МэВ^2 и $\mathrm{cov}(\hat{a}, \hat{b}) = -4$ канал^2/МэВ. Новое независимое измерение дало $y_* = 202$ канала с ошибкой $s_* = 2$ канала.

1. Найдите $\widehat{E} = (y_* - \hat{a})/\hat{b}$.
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.

РАЗДЕЛ КНИГИ

III. Статистический вывод

Главы

Глава 13

Качество согласия и статистическая значимость

Содержание

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 p-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

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

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 $\chi^2_{\min} = 12$. Во второй — $N = 1002$ точки и $\chi^2_{\min} = 1200$. Числа степеней свободы равны соответственно 10 и 1000. В обоих случаях

$$\frac{\chi^2_{\min}}{\nu} = 1.2. \tag{13.2}$$

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 $\chi^2_{\min}$ 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

$$r_i = y_i - \hat a - \hat b E_i, \qquad u_i = \frac{r_i}{\sigma_i}, \qquad \chi^2_{\min} = \sum_{i=1}^{N} u_i^2. \tag{13.3}$$

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

$$\sum_i \frac{r_i}{\sigma_i^2} = 0, \qquad \sum_i \frac{E_i r_i}{\sigma_i^2} = 0. \tag{13.4}$$

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 $\chi^2_{\min}$ — **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 $\chi^2_{\min}$ 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:

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

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

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

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

$$T = \sum_{i=1}^{\nu} G_i^2, \qquad T \sim \chi^2_\nu. \tag{13.7}$$

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

$$e^{-r^2/2} r^{\nu-1}\, dr.$$

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

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

Гамма-функция определяется интегралом

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

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 $\nu = 1$ 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 $t = 2u$:

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

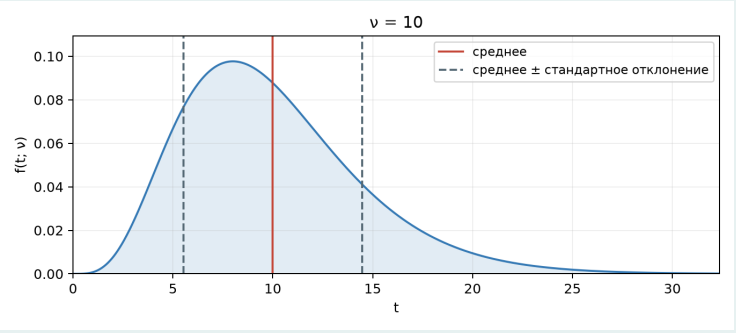
Отсюда

$$\operatorname{Var}(T) = 2\nu, \qquad \mathbb{E}[T/\nu] = 1, \qquad \sigma_{T/\nu} = \sqrt{\frac{2}{\nu}}. \tag{13.11}$$

Последняя формула объясняет проблему с правилом «χ^2 на степень свободы около единицы». При $\nu = 10$ стандартное отклонение отношения равно 0.447, при $\nu = 1000$ — 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 ν. Сплошная вертикальная линия показывает среднее, пунктир — расстояние $\sqrt{2\nu}$ от него. При $\nu = 1$ 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:

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

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:

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

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

$$\chi^2 = \sum_{i=1}^{N} g_i^2 \sim \chi^2_N. \tag{13.14}$$

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 m независимыми параметрами предсказания образуют m-мерное линейное подпространство. Здесь m — 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]

$$\chi^2_{\min} \sim \chi^2_\nu, \qquad \nu = N - m. \tag{13.15}$$

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

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

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

Это можно записать точным равенством

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

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

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

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

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

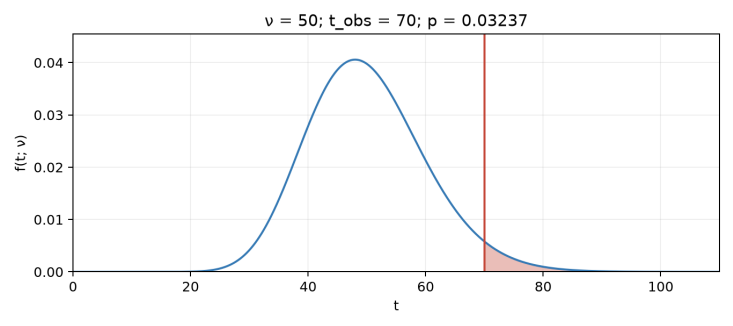
Спросим словами: если повторять эксперимент при верной H_0, в какой доле опытов получится сумма квадратов хотя бы такого размера?

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 $p \le \alpha$. 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 ν 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. p-значение как площадь хвоста
Для $\nu = 50$ и $t_{\text{obs}} = 70$ площадь правого хвоста равна $p \simeq 0.0324$. Закрашена вероятность получить $T \ge 70$ 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 p-значения?

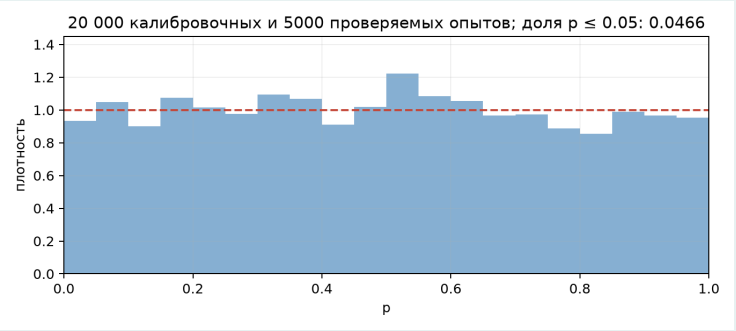
Представим много опытов, в которых H_0 верна. Для каждого вычислим T и его p-значение. Если функция распределения F_0 непрерывна, то $p = 1 - F_0(T)$. Преобразование переменной даёт

$$P_{H_0}(p \le u) = u, \qquad 0 \le u \le 1. \tag{13.19}$$

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:

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

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:

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

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:

$$\chi^2_{\min}/\mathrm{ndf} = 1200/1000, \qquad p = 1.226 \cdot 10^{-5}. \tag{13.22}$$

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

13.7. Что означает «число сигма»?

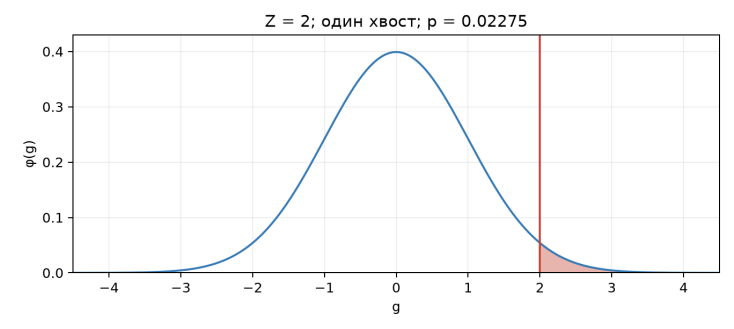
13.7.1. Один хвост

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

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

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 $Z \simeq 4.22$. 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	$2.867 \cdot 10^{-7}$

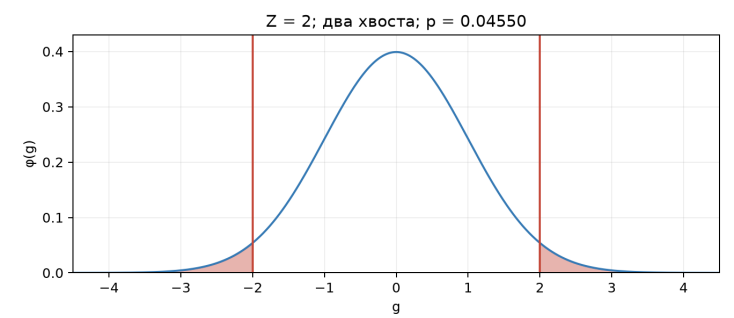
При поиске положительного сигнала обычно используют именно такое соглашение. Порог открытия 5σ соответствует правому хвосту из таблицы 13.2. Если $p > 1/2$, формула 13.23 даёт отрицательное Z. Такое число не означает свидетельства в пользу положительного сигнала.

13.7.2. Два хвоста

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

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

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

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

Его границы $\hat{\theta} \pm Z\sigma_\theta$ 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
$\hat{\theta} \pm \sigma_\theta$	$3.173 \cdot 10^{-1}$	1
$\hat{\theta} \pm 2\sigma_\theta$	$4.550 \cdot 10^{-2}$	4
$\hat{\theta} \pm 3\sigma_\theta$	$2.700 \cdot 10^{-3}$	9
$\hat{\theta} \pm 4\sigma_\theta$	$6.334 \cdot 10^{-5}$	16
$\hat{\theta} \pm 5\sigma_\theta$	$5.733 \cdot 10^{-7}$	25

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

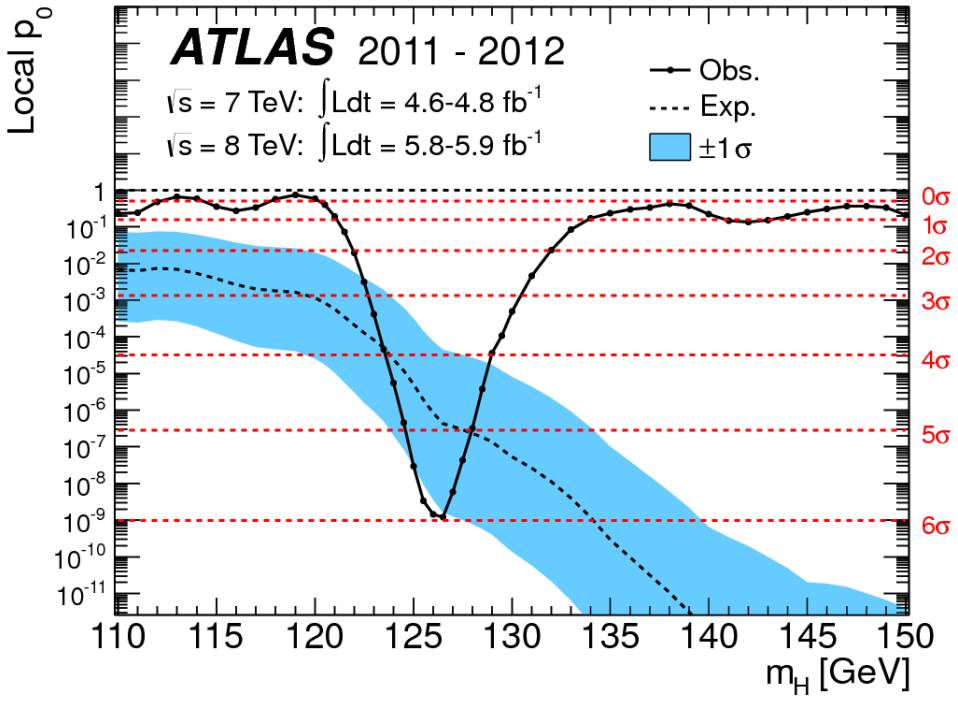


Рис. 13.1. *Локальное p-значение фоновой гипотезы в ATLAS. Сплошная чёрная кривая получена из данных; пунктир и голубая полоса относятся к ожидаемому результату при наличии бозона Хиггса Стандартной модели с пробной массой. Рисунок из статьи ATLAS [2].*

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.

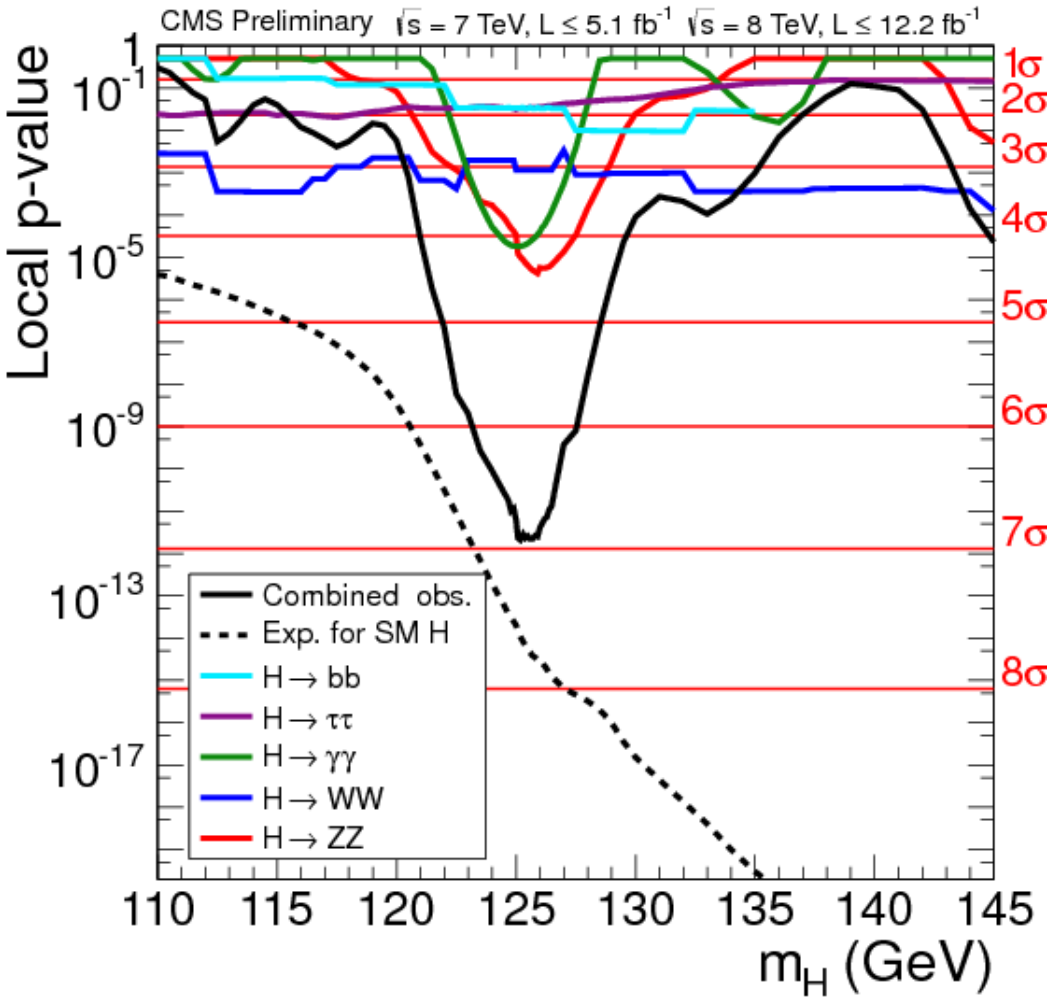


Рис. 13.2. *Локальные p-значения в предварительном результате CMS с данными при 7 и 8 ТэВ и светимостью до 12.2 фб$^{-1}$ при 8 ТэВ. Цветные кривые показывают отдельные каналы распада, чёрная — их объединение. Источник: CMS PAS HIG-12-045 [3].*

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

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

При верном фоне половина опытов имеет $Y \leq 0$ и даёт $q_0 = 0$. Положительная часть соответствует половине распределения χ^2_1. Для полученного $q_0 > 0$ имеем

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

Поэтому $q_0 = 25$ даёт односторонние 5σ. Для чистого χ^2_1 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:

$$N_i \sim \mathrm{Pois}(\mu_i), \qquad T_{\mathrm{P}} = \sum_{i=1}^{20} \frac{(N_i - \mu_i)^2}{\mu_i}. \tag{13.28}$$

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^2_{20}. 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 $\mathbb{E}[(N_i - \mu_i)^2] = \mu_i$. Но совпадения среднего недостаточно, чтобы совпало распределение. Используя центральный четвёртый момент Пуассона $\mathbb{E}[(N_i - \mu_i)^4] = \mu_i + 3\mu_i^2$, получаем

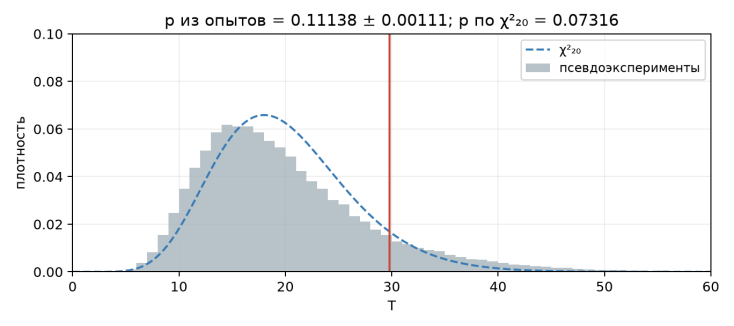
$$\mathrm{Var}(T_{\mathrm{P}}) = \sum_{i=1}^{20} \left(2 + \frac{1}{\mu_i}\right) \simeq 71.14. \tag{13.29}$$

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 $t_{\mathrm{obs}} = 29.8$ независимая проверка на миллионе опытов даёт $p \simeq 0.113$ 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 — $t_{\mathrm{obs}} = 29.8$. 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:

$$T_{\mathrm{L}} = 2 \sum_i \left[\mu_i - n_i + n_i \ln \frac{n_i}{\mu_i} \right]. \tag{13.30}$$

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

$$\hat{p} = \frac{k}{B}, \qquad \sigma_{\hat{p}} \simeq \sqrt{\frac{\hat{p}(1-\hat{p})}{B}}, \tag{13.31}$$

где B — число независимых опытов, k — число значений $T \geq t_{\mathrm{obs}}$. 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 $p \simeq 2.9 \cdot 10^{-7}$ 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 $N_i \sim \mathrm{Pois}(\mu_i)$. 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 $\mathbb{E}[(N_i - \mu_i)^4] = \mu_i + 3\mu_i^2$.
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 $\overline{Y}$, сумму квадратов относительно истинного μ и минимальную сумму квадратов относительно $\overline{Y}$; ошибки в знаменателях равны σ.
2. Проверьте точное разложение суммы квадратов на два слагаемых: минимум и $N(\overline{Y} - \mu)^2/\sigma^2$.
3. Сравните три распределения с χ^2_{12}, χ^2_{11} и χ^2_1. Проверьте их средние и дисперсии.
4. Объясните, почему остатки после подгонки не являются двенадцатью независимыми стандартными гауссовыми величинами.
5. Найдите p-значение для $\chi^2_{\min} = 12$. 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: $\mathbf{v} = \mathbf{w} - X(X^{\mathsf{T}}X)^{-1}X^{\mathsf{T}}\mathbf{w}$. 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 $\chi^2_{\min} = 12$ и p-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 $\chi^2_{\min}$.

Задача 4. Когда p-значения равномерны

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

1. Вычислите для каждого $p = P(\chi^2_{10} \geq T)$, 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^2_{10}. Какой ошибке в задании стандартных отклонений соответствует такая замена?
3. Сгенерируйте отдельную калибровочную выборку из 20 000 значений χ^2_{10}. Для исходных T вычислите ранговые p-значения $(k + 1)/(B + 1)$, где k — число калибровочных значений не меньше T, $B = 20000$.
4. Повторите опыт для $N \sim \mathrm{Pois}(2)$, 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 $p=0.05$, $2.7\cdot10^{-3}$ и $2.87\cdot10^{-7}$.
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, $s\geq0$, выведите $q_0=\max(0,Y/\sigma)^2$ для проверки $s=0$. Найдите вероятность $q_0\geq25$ при фоне. Сравните с $P(\chi^2_1\geq25)$.
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.

О книге

Статистическая процедура входит в определение экспериментального результата. Книга ведёт от вероятностной модели к физическому выводу и показывает, как оценки, интервалы и проверки гипотез меняются при конечной статистике, фоне и систематических неопределённостях.

МАРШРУТ ЧИТАТЕЛЯ

Вероятностный язык
Распределения, моменты, ЦПТ и границы гауссовости.

Моделирование
Распространение ошибок, корреляции и Монте-Карло.

Оценивание
Правдоподобие, свойства оценок и профилирование.

Физический вывод
χ^2, качество согласия, значимость и систематики.

Для студентов старших курсов, аспирантов и исследователей.

ОБ АВТОРЕ

Дмитрий В. Наумов — доктор физико-математических наук, теоретик и экспериментатор, специалист в физике нейтрино, элементарных частиц и астрофизике; руководитель Нейтринной программы ОИЯИ. Лауреат Breakthrough Prize in Fundamental Physics (2016) и премии Европейского физического общества (2023) в составе Daya Bay.