

Глава 10

Свойства оценок и профильное правдоподобие

Содержание

10.1.	Физическая задача: время жизни частицы	130
10.2.	Свойства оценок	130
10.3.	Информация Фишера и граница Крамера–Рао	133
10.4.	Форма правдоподобия около максимума	134
10.5.	Возвращаемся к времени жизни	135
10.6.	Малые выборки и физические границы	136
10.7.	Несколько параметров	137
10.8.	Пример: реакторные нейтрино	139
10.9.	Правдоподобие в реальном анализе	141
10.10.	Проверка численной подгонки	143
10.11.	Задачи	144
	Литература	145

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:

$$\hat{\tau} = \frac{1}{N} \sum_{i=1}^N t_i = 2.0 \text{ мкс.} \quad (10.1)$$

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

$$\hat{\sigma}_{\hat{\tau}} = \frac{\hat{\tau}}{\sqrt{N}} \simeq 0.45 \text{ мкс.} \quad (10.2)$$

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**:

$$b(\theta) = \text{bias}(\hat{\theta}) = \mathbb{E}_{\theta}[\hat{\theta}] - \theta. \quad (10.3)$$

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 $\text{Var}_{\theta}(\hat{\theta})$.

Чтобы учесть и смещение, и разброс, введём **средний квадрат ошибки**, или MSE (*mean squared error*):

$$\text{MSE}_\theta(\hat{\theta}) = \mathbb{E}_\theta[(\hat{\theta} - \theta)^2]. \quad (10.4)$$

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

$$\boxed{\text{MSE}_\theta(\hat{\theta}) = \text{Var}_\theta(\hat{\theta}) + b^2(\theta)}. \quad (10.5)$$

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:

$$\hat{\sigma}_{\text{ММП}}^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2, \quad s^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2. \quad (10.6)$$

Первая получается максимизацией правдоподобия. Вторая несмещённая. Их средние равны

$$\mathbb{E}[\hat{\sigma}_{\text{ММП}}^2] = \frac{N-1}{N} \sigma^2, \quad \mathbb{E}[s^2] = \sigma^2. \quad (10.7)$$

Но у s^2 больше разброс: каждый результат первой оценки умножен на $N/(N-1)$. Для гауссовой выборки точный расчёт даёт

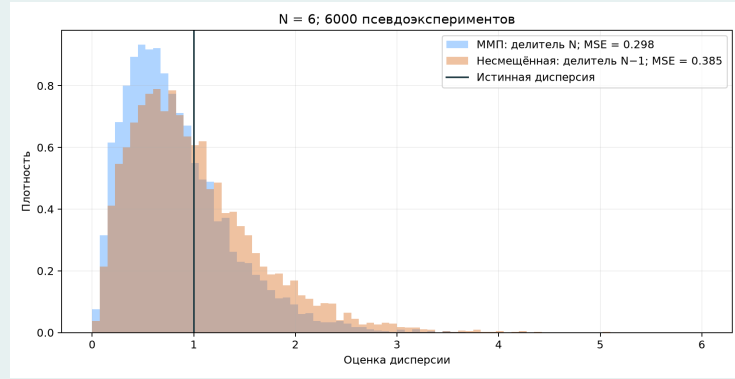
$$\text{Var}(\hat{\sigma}_{\text{ММП}}^2) = \frac{2(N-1)}{N^2} \sigma^4, \quad \text{Var}(s^2) = \frac{2}{N-1} \sigma^4. \quad (10.8)$$

Подставим эти выражения в формулу 10.5:

$$\text{MSE}(\hat{\sigma}_{\text{ММП}}^2) = \frac{2N-1}{N^2} \sigma^4, \quad \text{MSE}(s^2) = \frac{2}{N-1} \sigma^4. \quad (10.9)$$

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**:

$$P_{\theta}(|\hat{\theta}_N - \theta| > \varepsilon) \xrightarrow{N \rightarrow \infty} 0 \quad \text{при любом } \varepsilon > 0. \quad (10.10)$$

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,

$$P_{\theta}(|\hat{\theta}_N - \theta| > \varepsilon) \leq \frac{\text{MSE}_{\theta}(\hat{\theta}_N)}{\varepsilon^2}. \quad (10.11)$$

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

$$\begin{aligned} \text{Var}(\bar{X}) &= \frac{\sigma^2}{N}, \\ \text{Var}(\text{median } X_i) &\approx \frac{\pi}{2} \frac{\sigma^2}{N}, \\ \text{Var}\left(\frac{X_1 + X_N}{2}\right) &= \frac{\sigma^2}{2}, \quad N \geq 2. \end{aligned} \quad (10.12)$$

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*):

$$U(\theta) = \frac{\partial \ell(\theta)}{\partial \theta}. \quad (10.13)$$

До подстановки данных $U(\theta)$ является случайной величиной. Для информации Фишера существуют две равносильные записи:

$$I(\theta) = \mathbb{E}_\theta[U^2(\theta)] = -\mathbb{E}_\theta\left[\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right]. \quad (10.14)$$

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 $u = \partial_\theta \ln f(x | \theta)$. Дифференцируя условие нормировки, получаем

$$0 = \frac{\partial}{\partial \theta} \int f(x | \theta) dx = \int f(x | \theta) u dx = \mathbb{E}_\theta[u]. \quad (10.15)$$

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:

$$I_N(\theta) = \sum_{i=1}^N I_i(\theta). \quad (10.16)$$

Для одинаково распределённых наблюдений $I_N = NI_1$. 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 $E_\theta[T] = \theta$. Если производную снова можно перенести под интеграл, получаем

$$1 = E_\theta[TU] = E_\theta[(T - \theta)U].$$

По неравенству Коши–Буняковского квадрат среднего произведения не больше произведения средних квадратов. Следовательно,

$$1 \leq \text{Var}_\theta(T) I(\theta), \quad \boxed{\text{Var}_\theta(T) \geq \frac{1}{I(\theta)}}. \quad (10.17)$$

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 $E_\theta[T] = \theta$.

10.3.3. Пример: гауссово среднее

Для одного гауссова измерения с известной σ скор-функция равна

$$u_\mu(X) = \frac{X - \mu}{\sigma^2}, \quad I_1(\mu) = \frac{E[(X - \mu)^2]}{\sigma^4} = \frac{1}{\sigma^2}. \quad (10.18)$$

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:

$$\ell(\theta) \simeq \ell(\hat{\theta}) - \frac{1}{2} J(\hat{\theta})(\theta - \hat{\theta})^2, \quad J(\hat{\theta}) = -\ell''(\hat{\theta}). \quad (10.19)$$

Величина $J(\hat{\theta})$ определяется конкретной выборкой. При регулярной модели и достаточно большом N обратная кривизна даёт оценку дисперсии ММП:

$$\hat{\sigma}_{\hat{\theta}}^2 \simeq \frac{1}{J(\hat{\theta})}. \quad (10.20)$$

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:

$$\lambda(\theta) = \frac{L(\theta)}{L(\hat{\theta})}. \quad (10.21)$$

Теперь нормировка сократилась, а максимум равен единице. Для вычислений удобнее пользоваться логарифмом:

$$q(\theta) = -2 \ln \lambda(\theta) = -2[\ell(\theta) - \ell(\hat{\theta})]. \quad (10.22)$$

В максимуме $q = 0$, в остальных точках $q \geq 0$. В квадратичном приближении

$$q(\theta) \simeq \frac{(\theta - \hat{\theta})^2}{\hat{\sigma}_{\hat{\theta}}^2}. \quad (10.23)$$

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

$$q(\mu) = \frac{(\mu - \bar{x})^2}{\sigma^2/N}. \quad (10.24)$$

Здесь интервал $q(\mu) \leq 1$ имеет покрытие около 68.3%: при повторении эксперимента такая процедура покрывает 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

$$f(t | \tau) = \frac{1}{\tau} e^{-t/\tau}, \quad t \geq 0. \quad (10.25)$$

Для N независимых времён

$$L(\tau) = \tau^{-N} \exp\left(-\frac{1}{\tau} \sum_i t_i\right), \quad \ell(\tau) = -N \ln \tau - \frac{1}{\tau} \sum_i t_i. \quad (10.26)$$

Как обычно, в логарифме опущена константа, не зависящая от параметра. Условие максимума имеет вид

$$\frac{\partial \ell}{\partial \tau} = -\frac{N}{\tau} + \frac{\sum_i t_i}{\tau^2} = 0, \quad \hat{\tau} = \bar{t}. \quad (10.27)$$

Так мы получили формулу, с которой начали главу. У экспоненциальной величины $E[T_i] = \tau$ и $\text{Var}(T_i) = \tau^2$, поэтому

$$E[\hat{\tau}] = \tau, \quad \text{Var}(\hat{\tau}) = \frac{\tau^2}{N}. \quad (10.28)$$

Оценка несмещённая и состоятельная. Информация Фишера равна

$$I_N(\tau) = -E_{\tau} \left[\frac{N}{\tau^2} - \frac{2 \sum_i T_i}{\tau^3} \right] = \frac{N}{\tau^2}. \quad (10.29)$$

Дисперсия 10.28 достигает границы Крамера–Рао при любом N . Среди несмещённых оценок сделать её меньше в этой модели нельзя.

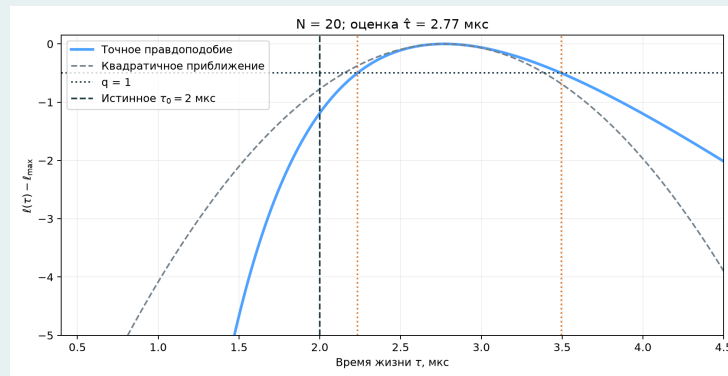
А вот распределение оценки при малом N ещё не гауссово: сумма экспоненциальных времён имеет гамма-распределение. Эффективность оценки и гауссова форма её распределения — разные свойства. Само правдоподобие также отличается от гауссова. Для него получаем точное выражение:

$$q(\tau) = 2N \left[\ln \frac{\tau}{\hat{\tau}} + \frac{\hat{\tau}}{\tau} - 1 \right]. \quad (10.30)$$

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

$$q(\mu) = 2 \left[\mu - n + n \ln \frac{n}{\mu} \right]. \quad (10.31)$$

При $n > 0$ функция определена для $\mu > 0$. При $n = 0$ используем предел $n \ln(n/\mu) = 0$, и формула упрощается до

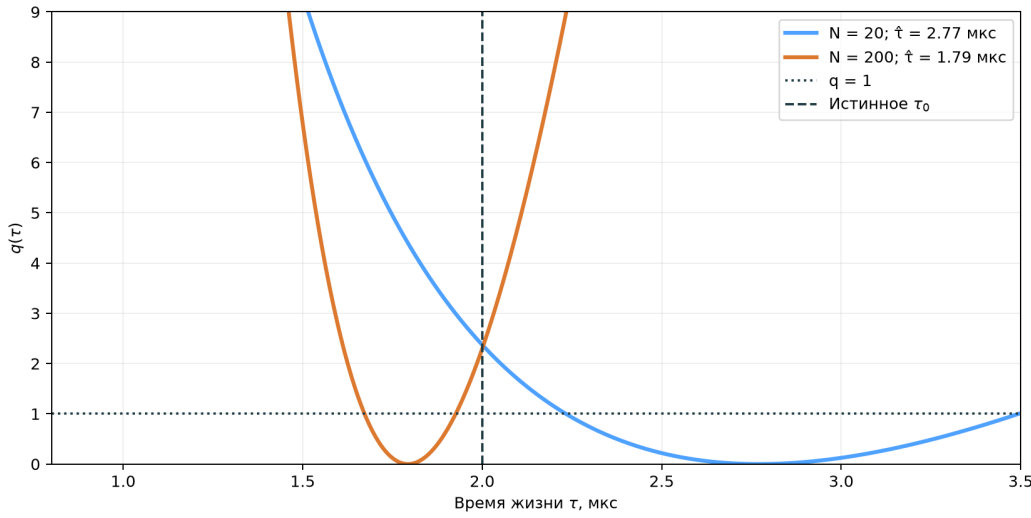


Рис. 10.1. Отношение правдоподобий для первых 20 и всех 200 времён одной последовательности распадов при $\tau_0 = 2$ мкс. Горизонтальная линия задаёт $q = 1$, вертикальная — истинное время жизни.

$$q(\mu) = 2\mu, \quad \mu \geq 0. \quad (10.32)$$

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:

$$(\hat{\theta}, \hat{\eta}) = \arg \max_{\theta, \eta} L(\theta, \eta). \quad (10.33)$$

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:

$$\hat{\eta}(\theta) = \arg \max_{\eta} L(\theta, \eta). \quad (10.34)$$

Двойная шляпка обозначает условный максимум при выбранном θ . Повторяя эту операцию для каждого θ , получим **профильное правдоподобие**:

$$\begin{aligned} L_p(\theta) &= L(\theta, \hat{\eta}(\theta)), \\ q_p(\theta) &= -2 \ln \frac{L_p(\theta)}{L(\hat{\theta}, \hat{\eta})}. \end{aligned} \quad (10.35)$$

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

$$q(\theta, \eta) = \frac{\theta^2 - 2\rho\theta\eta + \eta^2}{1 - \rho^2} = \theta^2 + \frac{(\eta - \rho\theta)^2}{1 - \rho^2}. \quad (10.36)$$

Вторая запись позволяет найти минимум по η . При фиксированном θ он достигается в точке

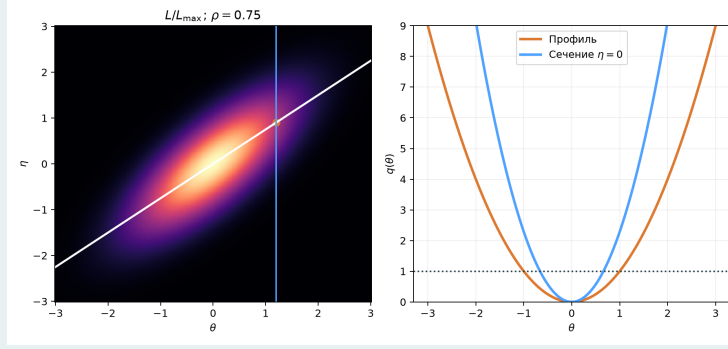
$$\hat{\eta}(\theta) = \rho\theta, \quad q_p(\theta) = \theta^2. \quad (10.37)$$

А сечение при $\eta = 0$ равно

$$q(\theta, 0) = \frac{\theta^2}{1 - \rho^2}. \quad (10.38)$$

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 $q_p(\theta)$ и сечение $q(\theta, 0)$. Белая линия показывает $\hat{\eta}(\theta) = \rho\theta$.



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:

$$P_{ee}(E; A, \Delta m^2) = 1 - A \sin^2\left(1.267 \Delta m^2 \frac{L}{E}\right). \quad (10.39)$$

В этой записи Δm^2 выражается в эВ², 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

$$\mu_i(A, \Delta m^2) = C \int_{E_i^-}^{E_i^+} \phi(E) \sigma(E) P_{ee}(E; A, \Delta m^2) dE. \quad (10.40)$$

В апплете интеграл приближён значением подынтегральной функции в центре бина, умноженным на его ширину:

$$\mu_i \simeq C \phi(E_i) \sigma(E_i) P_{ee}(E_i; A, \Delta m^2) \Delta E_i. \quad (10.41)$$

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 $n_i \sim \text{Pois}(\mu_i)$. Тогда с точностью до константы

$$\ell(A, \Delta m^2) = \sum_i [n_i \ln \mu_i(A, \Delta m^2) - \mu_i(A, \Delta m^2)]. \quad (10.42)$$

Эту функцию и предстоит максимизировать.

10.8.2. Спектр источника и сечение

Используем параметризацию спектров Хубера и Мюллера [3, 4]:

$$\phi(E) = \sum_k f_k \exp \left[\sum_{p=0}^5 a_{kp} (E/\text{МэВ})^p \right]. \quad (10.43)$$

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 $\bar{\nu}_e + p \rightarrow e^+ + n$. В приближении без отдачи нуклонов энергетическая зависимость сечения имеет вид

$$\sigma(E) \propto E_e p_e, \quad E_e = E - 1.293 \text{ МэВ}, \quad p_e = \sqrt{E_e^2 - m_e^2}, \quad (10.44)$$

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 .

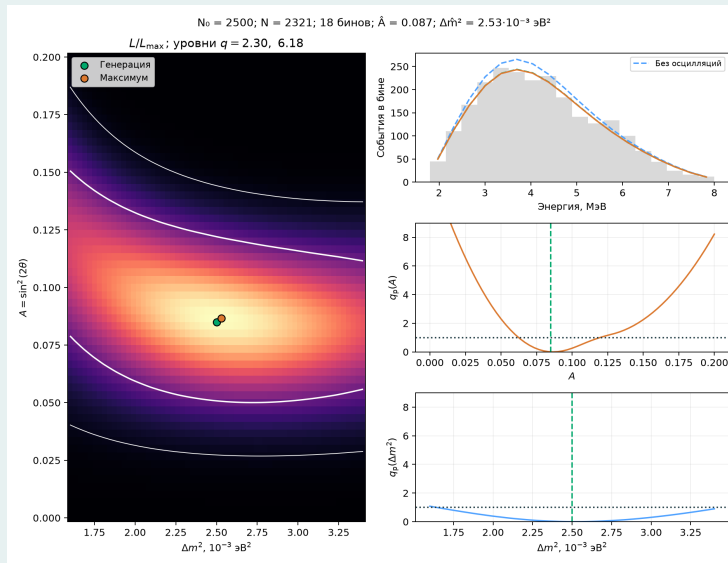
10.8.3. Генерация и подгонка

Сначала зададим истинные A_0 и Δm_0^2 , 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:

$$\begin{aligned} q(A, \Delta m^2) &= -2[\ell(A, \Delta m^2) - \ell_{\max}], \\ q_p(A) &= \min_{\Delta m^2} q(A, \Delta m^2), \\ q_p(\Delta m^2) &= \min_A q(A, \Delta m^2). \end{aligned} \quad (10.45)$$

Здесь оба параметра представляют физический интерес. Но при построении одномерного профиля одного из них второй играет роль мешающего.



Интерактив. Подгонка реакторного спектра

Псевдоэксперимент при $N_0 = 2500$, $A_0 = 0.085$, $\Delta m_0^2 = 2.5 \cdot 10^{-3} \text{ эВ}^2$ 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 $q_p = 1$ 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 $A \leq 0.20$ и $1.6 \leq \Delta m^2 / (10^{-3} \text{ эВ}^2) \leq 3.4$ 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

$$L_{\text{полн}}(\theta, \eta) = L_{\text{осн}}(\theta, \eta) \exp \left[-\frac{(a - \eta)^2}{2\sigma_a^2} \right] \times \text{const.} \quad (10.46)$$

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 $f(x | \theta)$ — нормированная плотность признака одного события, а $\nu(\theta)$ — ожидаемое число зарегистрированных событий. При пуассоновском счёте и независимых событиях **расширенное правдоподобие** равно

$$L_{\text{расш}}(\theta) = \frac{e^{-\nu} \nu^N}{N!} \prod_{i=1}^N f(x_i | \theta). \quad (10.47)$$

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:

$$L_{\text{расш}}(s, b) = \frac{e^{-(s+b)}}{N!} \prod_{i=1}^N [s f_s(x_i) + b f_b(x_i)]. \quad (10.48)$$

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:

$$L(\theta) = \prod_j \frac{e^{-\mu_j(\theta)} \mu_j(\theta)^{n_j}}{n_j!}. \quad (10.49)$$

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

$$\hat{\phi}_{\text{ММП}} = g(\hat{\theta}_{\text{ММП}}). \quad (10.50)$$

Это свойство называется **инвариантностью ММП**. Например, для скорости распада $\lambda = 1/\tau$ получаем $\hat{\lambda} = 1/\hat{\tau}$. Якобиан здесь не появляется: мы меняем аргумент функции правдоподобия. Преобразование плотности случайной величины — другая операция.

Несмещённость при нелинейной замене может потеряться, поскольку $E[g(\hat{\theta})]$ в общем случае не равно $g(E[\hat{\theta}])$. 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

$$X_i \sim N(\mu, \sigma^2),$$

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

$$\mathbb{E}[\hat{\sigma}_{\text{ММП}}^2] = \frac{N-1}{N} \sigma^2.$$

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

$$-2\ell(\theta, \eta) = \frac{\theta^2 - 2\rho\theta\eta + \eta^2}{1 - \rho^2} + \text{const}, \quad |\rho| < 1.$$

1. Найдите совместный максимум.
2. Для фиксированного θ найдите

$$\hat{\eta}(\theta).$$

3. Постройте профильную функцию $q_p(\theta)$.
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 бинировании.
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.