Вывод нормального распределения

Исторический контекст астрономические наблюдения

Ситуация: Астроном измеряет угловое положение звезды n раз, получая наблюдения:

$$y_1, y_2, ..., y_n$$

Каждое измерение содержит случайную ошибку.

Практический вопрос: Как наилучшим образом оценить истинное положение звезды μ?

Традиционный ответ: Взять среднее арифметическое:

$$\hat{\mu} = \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Фундаментальный вопрос:

При каком законе распределения ошибок измерений среднее арифметическое является **наилучшей** оценкой?

Аксиоматика Гаусса

Два ключевых допущения

Допущение $\boldsymbol{A_1}$: Среднее арифметическое — оптимальный способ оценки истинного значения.

$$\hat{\mu}_{\text{opt}} = \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Допущение $\boldsymbol{A_2}$: Наблюдения y_i — независимые одинаково распределённые случайные величины с плотностью $f(y \mid \mu)$, симметричной относительно μ:

$$f(\mu + t \mid \mu) = f(\mu - t \mid \mu) \quad \forall t$$

Метод максимального правдоподобия

Формализация

Оптимальная оценка $\hat{\mu}$ — это значение, максимизирующее совместную плотность:

$$\hat{\mu} = \arg\max_{\mu} f(y_1, y_2, ..., y_n \mid \mu)$$

Применение допущения A_2

Из независимости наблюдений:

$$f(y_1, ..., y_n \mid \mu) = \prod_{i=1}^{n} f(y_i \mid \mu)$$

Из допущения A_1: оптимум достигается при $\mu = \overline{y}$.

Связь A_1 и A_2:

$$\overline{y} = \arg\max_{\mu} \prod_{i=1}^{n} f(y_i \mid \mu)$$

Это функциональное уравнение на неизвестную плотность f!

Специальная выборка

Рассмотрим специальную выборку из $n+1$ наблюдения:

- n наблюдений со значением $\overline{y} - 1$
- 1 наблюдение со значением $\overline{y} + n$

Проверка: Среднее арифметическое:

$$\frac{n(\overline{y}-1) + (\overline{y}+n)}{n+1} = \frac{n\overline{y} - n + \overline{y} + n}{n+1} = \frac{(n+1)\overline{y}}{n+1} = \overline{y} \quad \checkmark$$

Функция правдоподобия

Для этой выборки:

$$L(\mu) = [f(\overline{y}-1 \mid \mu)]^n \cdot f(\overline{y}+n \mid \mu)$$

По A_1: максимум достигается при $\mu = \overline{y}$.

Переход к логарифму

Важное замечание: Нас интересует **где** достигается максимум, а не **чему равно** максимальное значение.

Поскольку ln — монотонная функция, $\arg\max L(\mu) = \arg\max \ln L(\mu)$.

$$\ell(\mu) = \ln L(\mu) = n \ln f(\overline{y}-1 \mid \mu) + \ln f(\overline{y}+n \mid \mu)$$

Условие первого порядка

Дифференцируем по μ и приравниваем к нулю:

$$\frac{d\ell}{d\mu} = n \cdot \frac{f'_\mu(\overline{y}-1 \mid \mu)}{f(\overline{y}-1 \mid \mu)} + \frac{f'_\mu(\overline{y}+n \mid \mu)}{f(\overline{y}+n \mid \mu)} = 0$$

При $\mu = \overline{y}$ (точка максимума):

$$n \cdot \frac{f'_\mu(\overline{y}-1 \mid \overline{y})}{f(\overline{y}-1 \mid \overline{y})} + \frac{f'_\mu(\overline{y}+n \mid \overline{y})}{f(\overline{y}+n \mid \overline{y})} = 0$$

Введение функции h

Определение: Введём функцию h, описывающую плотность отклонения от центра:

$$h(t) = f(\mu + t \mid \mu)$$

Интерпретация: $h(t)$ — плотность ошибки $t = y - \mu$.

Из симметрии (A_2): $h(t) = h(-t)$ — чётная функция.

Свойства h и h'

- $h(t)$ — **чётная**: $h(-t) = h(t)$
- $h'(t)$ — **нечётная**: $h'(-t) = -h'(t)$

Доказательство нечётности h':

Дифференцируем $h(-t) = h(t)$ по t:

$$-h'(-t) = h'(t) \quad \Rightarrow \quad h'(-t) = -h'(t)$$

Переписывание уравнения

Подставляем в условие первого порядка:

- $f(\overline{y} - 1 \mid \overline{y}) = h(-1)$
- $f(\overline{y} + n \mid \overline{y}) = h(n)$
- $f'_\mu(\overline{y} - 1 \mid \overline{y}) = h'(-1) = -h'(1)$ (нечётность!)
- $f'_\mu(\overline{y} + n \mid \overline{y}) = h'(n)$

Уравнение:

$$n \cdot \frac{-h'(1)}{h(-1)} + \frac{h'(n)}{h(n)} = 0$$

Используя чётность $h(-1) = h(1)$:

$$-n \cdot \frac{h'(1)}{h(1)} + \frac{h'(n)}{h(n)} = 0$$

Функциональное уравнение

Перепишем:

$$\frac{h'(n)}{h(n)} = n \cdot \frac{h'(1)}{h(1)}$$

Знак константы

Условие максимума: В точке $\mu = \overline{y}$ должен быть именно **максимум**, а не минимум.

Это требует $c < 0$.

Обозначение: $c = -\frac{1}{\sigma^2}$, где $\sigma > 0$.

Из вывода Гаусса получено:

Дифференциальное соотношение:

$$\frac{h'(x)}{h(x)} = -\frac{x}{\sigma^2}$$

Эквивалентные формы:

$$h'(x) = -\frac{x}{\sigma^2} \cdot h(x)$$

$$x \cdot h(x) = -\sigma^2 \cdot h'(x)$$

Ключевая идея: Мы **не решаем** это уравнение, а используем его для вычисления интегралов!

Граничные условия

На бесконечности для любой плотности вероятности:

$$\lim_{x \to \pm\infty} h(x) = 0$$

$$\lim_{x \to \pm\infty} x^k h(x) = 0 \quad \text{для любого } k$$

(экспоненциальное убывание h доминирует над полиномиальным ростом x^k)

Вычисление математического ожидания $\mathbb{E}[X]$

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x \cdot h(x) dx$$

Используем $x \cdot h(x) = -\sigma^2 \cdot h'(x)$:

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} (-\sigma^2) \cdot h'(x) dx = -\sigma^2 \int_{-\infty}^{+\infty} h'(x) dx$$

Вычисление

$$\int_{-\infty}^{+\infty} h'(x) dx = [h(x)]_{-\infty}^{+\infty} = h(+\infty) - h(-\infty) = 0 - 0 = 0$$

$$\mathbb{E}[X] = -\sigma^2 \cdot 0 = 0$$

Вычисление дисперсии $\mathbb{E}[X^2]$

$$\mathbb{E}[X^2] = \int_{-\infty}^{+\infty} x^2 \cdot h(x) dx$$

Перепишем $x^2 \cdot h(x) = x \cdot (x \cdot h(x))$ и используем $x \cdot h(x) = -\sigma^2 h'(x)$:

$$\mathbb{E}[X^2] = \int_{-\infty}^{+\infty} x \cdot (-\sigma^2 h'(x)) dx = -\sigma^2 \int_{-\infty}^{+\infty} x \cdot h'(x) dx$$

Интегрирование по частям

$$\int x \cdot h'(x)dx = \underbrace{x \cdot h(x)}_{u \cdot v} - \int \underbrace{h(x)}_{v} \underbrace{dx}_{du}$$

Вычисляем:

$$\int_{-\infty}^{+\infty} x \cdot h'(x)dx = [x \cdot h(x)]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} h(x)dx$$

- Первое слагаемое: $[x \cdot h(x)]_{-\infty}^{+\infty} = 0 - 0 = 0$ (граничные условия)
- Второе слагаемое: $\int_{-\infty}^{+\infty} h(x)dx = 1$ (нормировка)

$$\int_{-\infty}^{+\infty} x \cdot h'(x)dx = 0 - 1 = -1$$

$$\mathbb{E}[X^2] = -\sigma^2 \cdot (-1) = \sigma^2$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \sigma^2 - 0 = \sigma^2$$

Параметр σ^2 — это в точности дисперсия!

Анализ размерностей

Размерность плотности вероятности

Физический смысл плотности:

$$P(x < X < x + dx) = h(x)dx$$

Размерности:

- P — безразмерная величина (вероятность)
- dx — имеет размерность $[x]$ (например, градусы)
- Следовательно: $[h(x)] = \frac{1}{[x]}$ (обратная размерность!)

Анализ экспоненты

Аргумент экспоненты должен быть безразмерным:

$$\left[\frac{x^2}{\sigma^2}\right] = \frac{[x]^2}{[\sigma]^2} = 1 \quad \Rightarrow \quad [\sigma] = [x]$$

Размерность σ совпадает с размерностью x (это стандартное отклонение).

Размерность константы A

$$[h(x)] = [A] \cdot [\exp(...)] = [A] \cdot 1 = [A]$$

Следовательно:

$$[A] = \frac{1}{[x]} = \frac{1}{[\sigma]}$$

Вывод из размерностей:

$$A = \frac{\text{const}}{\sigma}$$

где const — безразмерная численная константа.

Условие нормировки

$$\int_{-\infty}^{+\infty} h(x)dx = 1$$

$$A \int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = 1$$

Вычисление интеграла Гаусса

Замена $u = \frac{x}{\sigma\sqrt{2}}$, $dx = \sigma\sqrt{2}du$:

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \sigma\sqrt{2} \int_{-\infty}^{+\infty} e^{-u^2} du = \sigma\sqrt{2} \cdot \sqrt{\pi} = \sigma\sqrt{2\pi}$$

Константа нормировки

$$A \cdot \sigma\sqrt{2\pi} = 1 \quad \Rightarrow \quad A = \frac{1}{\sigma\sqrt{2\pi}}$$

Плотность нормального распределения:

$$h(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

Или в терминах исходной плотности:

$$f(y \mid \mu) = h(y - \mu) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

Красота метода:

Мы получили все моменты нормального распределения, **не зная явного вида** плотности $h(x)$!

Достаточно только соотношения $\frac{h'}{h} = -\frac{x}{\sigma^2}$ и граничных условий.

Это пример того, как **дифференциальное уравнение** содержит больше информации, чем кажется на первый взгляд.

Рекуррентная формула для моментов

Вычисление $\mathbb{E}[X^{2k}]$

$$\mathbb{E}[X^{2k}] = \int_{-\infty}^{+\infty} x^{2k} \cdot h(x)dx$$

Перепишем:

$$x^{2k} \cdot h(x) = x^{2k-1} \cdot (x \cdot h(x)) = x^{2k-1} \cdot \left(-\sigma^2 h'(x)\right)$$

$$\mathbb{E}[X^{2k}] = -\sigma^2 \int_{-\infty}^{+\infty} x^{2k-1} \cdot h'(x)dx$$

Интегрируем по частям

$$\int x^{2k-1} h'(x)dx = \left[x^{2k-1} h(x)\right] - \int (2k-1)x^{2k-2} h(x)dx$$

Вычисляем:

- $\left[x^{2k-1} h(x)\right]_{-\infty}^{+\infty} = 0$ (граничные условия)
- $\int x^{2k-2} h(x)dx = \mathbb{E}[X^{2k-2}]$

$$\int_{-\infty}^{+\infty} x^{2k-1} h'(x)dx = 0 - (2k-1)\mathbb{E}[X^{2k-2}] = -(2k-1)\mathbb{E}[X^{2k-2}]$$

Рекуррентное соотношение

$$\mathbb{E}[X^{2k}] = -\sigma^2 \cdot \left(-(2k-1)\mathbb{E}[X^{2k-2}]\right) = \sigma^2 (2k-1)\mathbb{E}[X^{2k-2}]$$

Рекуррентная формула:

$$\mathbb{E}[X^{2k}] = (2k-1) \cdot \sigma^2 \cdot \mathbb{E}[X^{2k-2}]$$

Начальное условие: $\mathbb{E}[X^0] = 1$

Первые моменты

k	Рекуррентность	$\mathbb{E}[X^{2k}]$	Коэффициент
0	—	1	1
1	$1 \cdot \sigma^2 \cdot 1$	σ^2	1
2	$3 \cdot \sigma^2 \cdot \sigma^2$	$3\sigma^4$	$3 = 1 \cdot 3$
3	$5 \cdot \sigma^2 \cdot 3\sigma^4$	$15\sigma^6$	$15 = 1 \cdot 3 \cdot 5$
4	$7 \cdot \sigma^2 \cdot 15\sigma^6$	$105\sigma^8$	$105 = 1 \cdot 3 \cdot 5 \cdot 7$
5	$9 \cdot \sigma^2 \cdot 105\sigma^8$	$945\sigma^{10}$	$945 = 1 \cdot 3 \cdot 5 \cdot 7 \cdot 9$

Общая формула

$$\mathbb{E}[X^{2k}] = (2k-1)!! \cdot \sigma^{2k}$$

где $(2k-1)!! = 1 \cdot 3 \cdot 5 \cdot ... \cdot (2k-1)$ — двойной факториал.

Альтернативная форма:

$$\mathbb{E}[X^{2k}] = \frac{(2k)!}{2^k \cdot k!} \cdot \sigma^{2k}$$

Нечётные моменты

$$\mathbb{E}[X^{2k+1}] = 0$$

Причина: $x^{2k+1} \cdot h(x)$ — нечётная функция (произведение нечётной x^{2k+1} и чётной $h(x)$).

Интеграл нечётной функции по симметричному промежутку равен нулю.

Вычисление $\mathbb{E}[X^{2026}]$

Определение k

$$2026 = 2 \cdot 1013 \quad \Rightarrow \quad k = 1013$$

Применение формулы

$$\mathbb{E}[X^{2026}] = (2 \cdot 1013 - 1)!! \cdot \sigma^{2026} = 2025!! \cdot \sigma^{2026}$$

Ответ:

$$\mathbb{E}[X^{2026}] = 2025!! \cdot \sigma^{2026} = (1 \cdot 3 \cdot 5 \cdot ... \cdot 2025) \cdot \sigma^{2026}$$

$$= \prod_{j=1}^{1013} (2j-1) \cdot \sigma^{2026}$$

Численная оценка (при $\sigma = 1$)

$$2025!! = \prod_{j=1}^{1013} (2j-1)$$

Используем формулу Стирлинга для двойного факториала:

$$\ln(2k-1)!! \approx k\ln(2k) - k + \frac{1}{2}\ln(\pi k)$$

При $k = 1013$:

$$\ln(2025!!) \approx 1013\ln(2026) - 1013 + \frac{1}{2}\ln(1013\pi) \approx 5906$$

$$2025!! \approx e^{5906} \approx 10^{2565}$$

Историческое значение:

Гаусс показал: если мы **постулируем** оптимальность среднего арифметического как оценки, то закон распределения ошибок **однозначно определяется** — это нормальное (гауссово) распределение.

Это замечательный пример «обратной задачи» в статистике: не «какая оценка оптимальна для данного распределения?», а «какое распределение делает данную оценку оптимальной?»