Математическая статистика.

Итоги:

- LMS провалился !!
- формула = $\underbrace{0{,}2\cdot Д3 + 0{,}25 КР1 + 0{,}25 КР2}$ + 0,3·экзам.

→ округл. [1–8] ?

→ отсылка к новой уе → да !!

→ отсылка на русск → нет !!

xkcd of statistics

В теор. делали предп о вер. P(...)

⇓

точно считали P(...), E(...), H(...), Var(...)

В статист. Есть неизвестный параметр θ
предположения о вер P(...)

⇓

оценить P(...), E(...), H(...), Var(...)

заранее согласны на ошибку.

mathematics

statistics

---

① → Метод моментов

② → Метод максимального правдоподобия.

---

① Метод моментов.

не знаем: θ – неизвестный параметр $\theta \in \mathbb{R}^1$ $\theta \in \Theta$

константа.

наблюдаем: $y_1, y_2, \ldots\ldots y_n$ (СВ)

$F(y_1, y_2 \ldots\ldots y_n) \leftarrow$ зависит от θ.
(и от $y_1 \ldots\ldots y_n$, кон.-ко)

ф. расп.

Цель: оценить θ. $\boxed{\text{Найти } \hat{\theta}.}$ $\boxed{\hat{\theta} \neq \theta}$

Предположим: $y_1, y_2 \ldots\ldots y_n$ — независимы, одинаково распр-ны.

$F(y_i / \theta) \leftarrow$ ф. расп. каждого y_i зав. от θ.

Посмотрим на З.Б.Ч.

$$\operatorname*{plim}_{n\to\infty} \frac{y_1 + \ldots\ldots + y_n}{n} = E(y_1)$$

в быту „среднее"

$\frac{y_1 + \ldots y_n}{n}$ — среднее выборочное
sample mean

$E(y_1)$ — мат.-ое ожидание
expected value/
population mean

Из З.Б.Ч. следует идея $\boxed{\text{если } n \text{ велико}}$, то:

$$\frac{y_1 + \ldots + y_n}{n} \approx \underbrace{E(y_1)}_{} = h(\theta)$$

зависит от θ

Способ оценивания: Метод моментов

Найдите $\hat{\theta}$ решив уравнение

$$\frac{y_1 + \ldots\ldots + y_n}{n} = h(\theta) \quad \text{относ-но } \theta$$

Пример. $y_1, y_2, \ldots, y_n$ независимы, $Expo(\lambda)$

Выборка: $n=5$ $y_1=7$ $y_2=3$ $y_3=6$ $y_4=8$ $y_5=10$

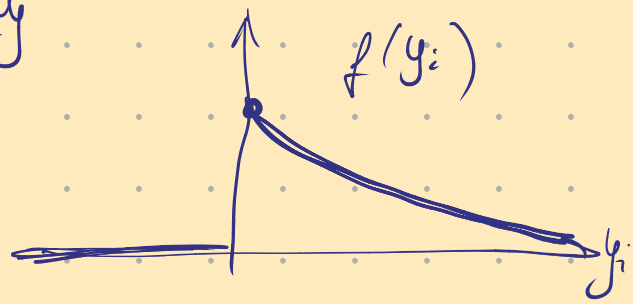
Оцените λ с помощью метода моментов.

$$\boxed{\hat{\lambda} \stackrel{?}{=}}$$

$$f(y_i) = \begin{cases} \lambda \cdot e^{-\lambda y_i}, & y_i \geq 0 \\ 0, & \text{иначе} \end{cases}$$

$$h(\theta) = E(y_1) = \frac{1}{\lambda}$$

$$\int_0^\infty y \cdot \lambda \cdot e^{-\lambda y} \, dy$$



Если n велико, то

$$\frac{y_1 + \ldots + y_n}{n} \approx \frac{1}{\lambda}$$

$$\frac{y_1 + \ldots + y_n}{n} = \frac{1}{\hat{\lambda}}$$

$\hat{\lambda}$ — оценка метода моментов.

$$\boxed{\hat{\lambda} = \frac{1}{\frac{y_1 + \ldots + y_n}{n}} = \frac{n}{y_1 + \ldots + y_n}}$$

$$\hat{\lambda} = \frac{5}{7+3+6+8+10}$$

$$\boxed{\hat{\lambda} = \frac{1}{\bar{y}}} \qquad \bar{y} = \frac{y_1 + \ldots + y_n}{n}$$

оценка метода моментов.

Проблемы? (1) h не зависит от неизв.-го пар-ра θ.

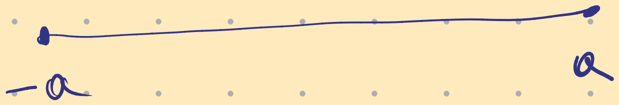
Пример. $y_1, \ldots, y_n$ — независимы $Unif[-a; a]$

наблюдаем

a — неизв. пар-р

Оцените a методом моментов.

$\underline{h(a) = E(y_1)} = 0$



$\boxed{\text{Тупик}}$: надо искать a из ур-ия

$$\frac{y_1+y_2+\dots+y_n}{n} = 0 \quad !!$$

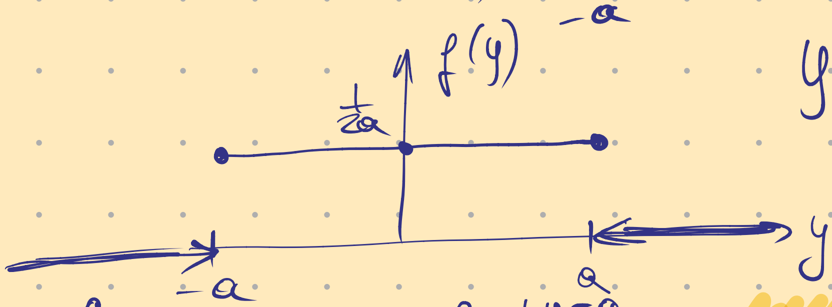
Решение: считать не $E(y_1)$, а $E(\quad)$

какая-то ф-ция от y_1, а не сам y_1

Простая идея: степень y_1 вторую, третью, ...

$$E(y_1^2) = \int_{-a}^{a} y^2 \cdot f(y)\,dy =$$

$y_i \sim Unif[-a;a]$



$$= \int_{-a}^{a} y^2 \cdot \frac{1}{2a}\,dy = \frac{y^3}{6a}\Big|_{y=-a}^{y=a} = \frac{1}{3}a^2$$

Преобразуем данные

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \to \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \qquad \boxed{w_i = y_i^2}$$

Используем метод моментов для преобр-ой переменной.

ЗБЧ: при $n >> 0$:

$$\underbrace{\frac{w_1+\dots w_n}{n}}_{\text{среднее выборочное}} \approx \underbrace{E(w_1)}_{\text{мат-ое ожидание}} = \frac{a^2}{3}$$

Метод моментов

$$\frac{w_1 + \ldots + w_n}{n} = \frac{\hat{a}^2}{3}$$

Решаем: $\boxed{a>0}$

$$\boxed{\hat{a} = \sqrt{3 \cdot \frac{w_1 + \ldots + w_n}{n}}}$$

от выбора преобразования зависит оценка.

$y_i \to w_i = y_i^2$

$y_i \to s_i = |y_i|$

$$E(s_i) = E(|y_i|) \overset{?}{=} \int_{-a}^{a} |y| \cdot f(y)\, dy$$

$y_i \sim Unif[-a; a]$

$-a \quad 0 \quad a$

$|y_i| = s_i \sim Unif[0; a]$

$0 \quad a$

$$E(s_i) = \frac{a}{2}$$

З.Б.Ч.

$$\frac{s_1 + s_2 + \ldots + s_n}{n} \approx \frac{a}{2}$$

Метод моментов:

$$\frac{s_1 + \ldots + s_n}{n} = \frac{\hat{a}}{2}$$

$$\boxed{\hat{a} = 2 \cdot \frac{|y_1| + |y_2| + \ldots + |y_n|}{n}}$$

Названия:

→ теоретич. момент (k)-го порядка $E(y_i^k)$

→ выборочный момент (k)-го порядка $\frac{y_1^k + y_2^k + \ldots + y_n^k}{n}$

Проблема ②: А если надо оценить несколько пар-ров?

Пример. $y_1, y_2, \dots, y_n \sim$ нерав $U[a;b]$

наблюдения — неизв. пар-ры.

ф-ции можно взять другие

$$\begin{cases} E(y_i) = \frac{a+b}{2} \\ E(y_i^2) = Var(y_i) + (E(y_i))^2 = \frac{(b-a)^2}{12} + \left(\frac{a+b}{2}\right)^2 \end{cases}$$

Метод моментов: $\hat{a}$ и $\hat{b}$ нах-ся как решение системы

$$\begin{cases} \frac{y_1 + \dots + y_n}{n} = \frac{\hat{a}+\hat{b}}{2} \\ \frac{y_1^2 + y_2^2 + \dots + y_n^2}{n} = \frac{(\hat{b}-\hat{a})^2}{12} + \frac{(\hat{a}+\hat{b})^2}{4} \end{cases}$$

$\to \hat{a}, \hat{b}$.

Метод максимального правдоподобия.

Идея: частые события случаются часто, а редкие - редко.

Способ: давайте считать что то, что произошло, это частое событие.

$\downarrow$

(формализация) $\hat{\theta} = \underset{\theta}{\arg\max}\, P(A|\theta)$ (A - произошедшее событие) $\leftarrow$ в дискретном

$\hat{\theta} = \underset{\theta}{\arg\max}\, f(y_1 \dots y_n|\theta) \leftarrow$ для величин с ф-цией плотности

Пример.

r	Караси	Щуки	Налимы
$P(R_i = r)$	θ	$\theta/2$	$1-\frac{3}{2}\theta$

R_i ~ независимые

Наблюд.: $n=4$

$R_1 =$ Кар
$R_2 =$ Щук
$R_3 =$ Нал
$R_4 =$ Кар

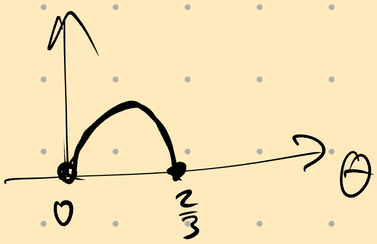
$\hat{\theta}$ с помощью макс. правдоп.

метод макс правд: то, что произошло, самое вероятное событие!

$$\boxed{P(R_1 = \text{Кар}, R_2 = \text{Щ}, R_3 = \text{Н}, R_4 = \text{К}) \to \max_{\theta}}$$

$$P(R_1 = \text{К}) \cdot P(R_2 = \text{Щ}) \cdot P(R_3 = \text{Н}) \cdot P(R_4 = \text{К}) =$$

$$= \theta \cdot \frac{\theta}{2} \cdot \theta \cdot \left(1 - \frac{3}{2}\theta\right) = \boxed{\frac{1}{2} \cdot \theta^3 \cdot \left(1 - \frac{3}{2}\theta\right)} \to \max_{\theta}$$



Технич. хитрость

ln

[на практике! перемножаем много мал. чисел]

$$\ell(\theta) = \ln P(R = (\text{К}, \text{Щ}, \text{Н}, \text{К}) \mid \theta) = \ln\frac{1}{2} + 3\ln\theta + \ln\left(1 - \frac{3}{2}\theta\right)$$

$$\ell'(\theta) = \frac{3}{\theta} - \frac{3}{2} \cdot \frac{1}{1 - \frac{3}{2}\theta}$$

[! надо убедиться, что это максимум]

$$\frac{3}{\hat{\theta}} = \frac{3}{2} \cdot \frac{1}{1 - \frac{3}{2}\hat{\theta}}$$

$$6 \cdot \left(1 - \frac{3}{2}\hat{\theta}\right) = 3\hat{\theta}$$

$$6 - 9\hat{\theta} = 3\hat{\theta}$$

$$\boxed{\hat{\theta} = \frac{6}{12} = 0.5}$$