Семинар 19

Неформ-но

Когда бутстрэп применим?

- Если при оценке пар-ра (возм-но преобраз-го)

ошибка $R = h(\hat{\theta}) - h(\theta)$
оцен-ия
- симм-на отн 0
- непрерывно (строгое мон-сть)

В частности бутстрэп точно работает для асимп-ки норм-х оценок.

Когда не работает бутстрэп?

→ min / max / края огр. распределения

Упр (с практич. ценностью)

Хотим реали-ть бутстрэп в „режиме реаль-ного времени"
знаем, что $n >> 0$. (оч. велико).

Хотим напротив каждого икс-го наблюдения написать, сколько раз его брать в бутстрэп выборку.

X_1	$Q_1=1$
X_2	$Q_2=2$
X_3	$Q_3=0$
$\vdots$	
X_n	$Q_n=2$

$n \to \infty$

а) Как распределены Q_i?

$P(Q_i = k)$?
$E(Q_i)$?
$Var\, Q_i$?

Q_i = кол-во раз, к-рое наблюдение номер i попадает в бутстрэп выборку.

б) $\mathrm{Cov}(Q_i, Q_j)$?   ($n \to \infty$)

(n велико)

догадаться? $Q_i \in \{0, 1, 2, 3, 4, \dots\ n\}$

- биноминальное ☺
- пуассон
- геом-ское
- отриц. бином.
- гипер-геом

?

$P(Q_i = k) = ?$ i-ое наблюдение при каждой попытке можем взять или не взять

i фиксир

$P(X_j^* = X_i) = \frac{1}{n}$

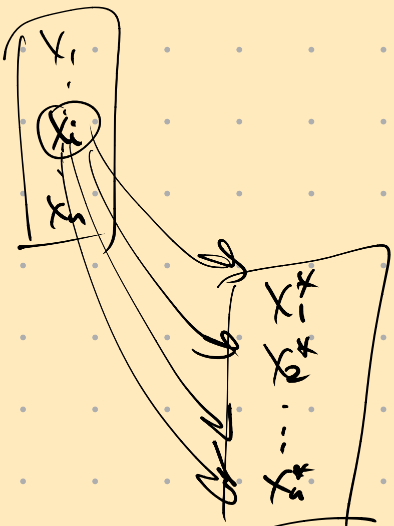


Схема Бернулли в чистом виде!

$Q_i \sim Bin\left(n, p = \frac{1}{n}\right)$

Если
$$p \to 0$$
$$n \to \infty$$
$$p \cdot n \to \lambda > 0$$
то
$$Bin(n, p) \xrightarrow{dist} Poiss(\lambda)$$

$$p \cdot n = \frac{1}{n} \cdot n = 1 \xrightarrow{!!} 1 = \lambda$$

$$Q_i \xrightarrow{dist} \text{Poiss}(\lambda = 1)$$

$$P(Q_i = k) = \exp(-\lambda)\cdot\frac{\lambda^k}{k!} \overset{\lambda=1}{=} \exp(-1)\frac{1}{k!} = \frac{\exp(-1)}{k!}$$

док-во основано на 2-м зам-м пределе

$\boxed{\text{например}}$

$$P(Q_i = 3) = C_n^3 \left(\frac{1}{n}\right)^3 \left(\frac{n-1}{n}\right)^{n-3} =$$

$$Q_i \sim Bin\left(n, \frac{1}{n}\right)$$

$$= C_n^3 \cdot \frac{(n-1)^{n-3}}{n^n} \cdot \frac{(n-1)^3}{(n-1)^3} =$$

$$= \frac{n\cdot(n-1)\cdot(n-2)}{3!\,(n-1)^3} \cdot \frac{(n-1)^n}{n^n} =$$

$$= \frac{n\,(n-1)\cdot(n-2)}{(n-1)\cdot(n-1)\cdot(n-1)\cdot 3!} \cdot \left(1 - \frac{1}{n}\right)^n \xrightarrow[n\to\infty]{} \frac{1}{3!}e^{-1}$$

$$\left(1 + \frac{k}{n}\right)^n \to e^k$$

a) $E(Q_i) = np = n\cdot\frac{1}{n} = 1 \to 1$

$$Var(Q_i) = np(1-p) = n\cdot\frac{1}{n}\cdot\left(1 - \frac{1}{n}\right) = 1 - \frac{1}{n} \xrightarrow{n\to\infty} 1$$

б) $Cov(Q_1, Q_2) =$

ков-ция легко работает с суммами, разложим Q_1 и Q_2 в суммы

$$= Cov(I_1 + I_2 + \ldots + I_n,\ J_1 + J_2 + \ldots + J_n) =$$

$I_i = \{ 1$, если $x_i^* = x_1$; 0

$J_i = \{ 1$, если $x_i^* = x_2$; 0

$$\mathrm{Cov}(I_1, J_3) = 0$$

разл.

$$= \mathrm{Cov}(I_1, J_1) + \mathrm{Cov}(I_2, J_2) + \ldots + \mathrm{Cov}(I_n, J_n) =$$

$$= n \cdot \mathrm{Cov}(I_1, J_1) = n \cdot \left(-\frac{1}{n^2}\right) = -\frac{1}{n}$$

	$J_1 = 0$	$J_1 = 1$
$I_1 = 0$	$\frac{n-2}{n}$	$1/n$
$I_1 = 1$	$1/n$	0

$X_1^* = X_2$

$X_1^* = X_1$

Выбираем из $X_1 \ldots X_n$ значение для X_1^*.

$$I_1 \cdot J_1 \equiv 0$$

$$\mathrm{Cov}(I_1, J_1) = E(I_1 \cdot J_1) - E(I_1) \cdot E(J_1) =$$

$$= 0 - \frac{1}{n} \cdot \frac{1}{n} = -\frac{1}{n^2}$$

$$\lim_{n \to \infty} \mathrm{Cov}(Q_1, Q_2) = \lim_{n \to \infty} \left(-\frac{1}{n}\right) = 0.$$

практич смысл этой задачи?

я заранее не знаю!

t

$x_1 \quad x_2 \quad x_3 \quad \ldots \quad x_n$

бут. выб. №1 $\quad Q_1 \quad Q_2 \quad Q_3 \quad \ldots \quad Q_n$

$n = 10^3$

$B = 10^4$

$Q_i \sim \mathrm{Poiss}(\lambda = 1)$

$S = X_1 \cdot Q_1 + X_2 \cdot Q_2 + X_3 \cdot Q_3 + \ldots$

В конце $\quad \bar{X}^* = \frac{S}{n}$

нау-ся   Пуассоновский бутстрэп.

Упр-пример.

$X_1, X_2, \ldots, X_n$ ~ незав. один. распр $N(\mu;1)$

А-приори знаю, что $\mu \in [0;\infty)$

а) $\hat{\mu}_{ML}$ (мет. максим-го правдоп.)

б) будет ли у наивного бутстрэпа фактич. вер-сть накрытия стремиться при $n \to \infty$ к номинальной?

а) $\max\limits_{\mu \in [0;\infty)} l(\mu)$

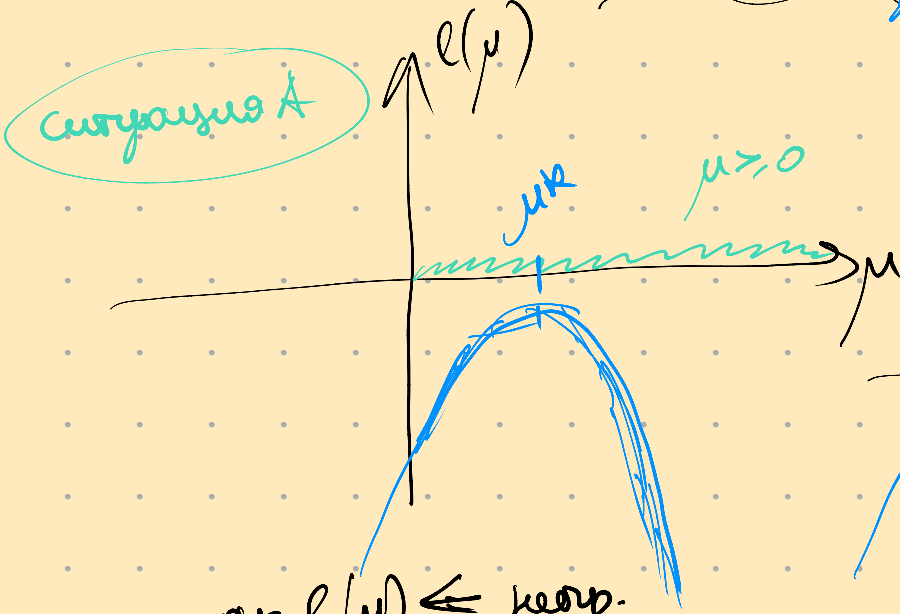
$$l(\mu) = \ln f(x_1, \ldots, x_n) = \ln f(x_1) + \ln f(x_2) + \ldots + \ln f(x_n)$$

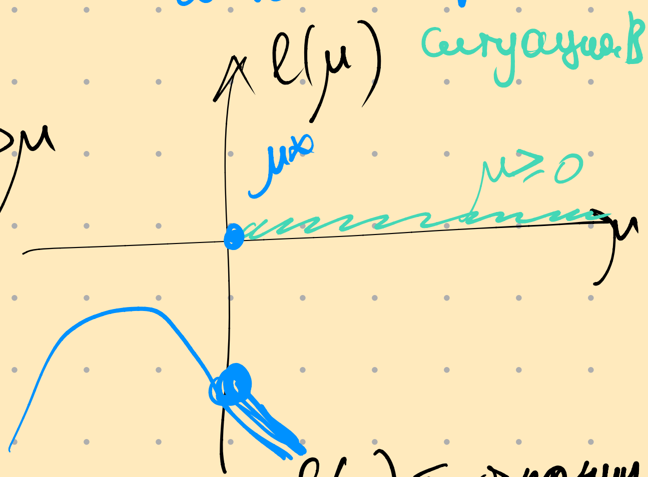
$X_i \sim N(\mu;1)$

$$f = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \qquad \sigma^2 = 1$$

$$\ln f(x_i) = -\frac{1}{2}(\ln 2\pi) - \frac{(x_i-\mu)^2}{2}$$

отн пер-ой μ — чистая парабола.





$\max\limits_{\mu \in [0;\infty)} l(\mu) \Leftarrow$ безр.

$l'_\mu = 0$

$\max\limits_{\mu \in [0;\infty)} l(\mu) \Leftarrow$ огранич.

решаем $\hat{\mu} = 0.$

$l'(\mu)=0$

$$\frac{\partial \sum (x_i-\mu)^2}{\partial \mu} = 0$$

$$\sum 2(x_i-\hat{\mu})\cdot(-1)=0.$$

$$\sum (x_i-\hat{\mu})=0$$

$$\sum x_i - \sum \hat{\mu} = 0$$

$$\sum x_i - n\hat{\mu} = 0$$

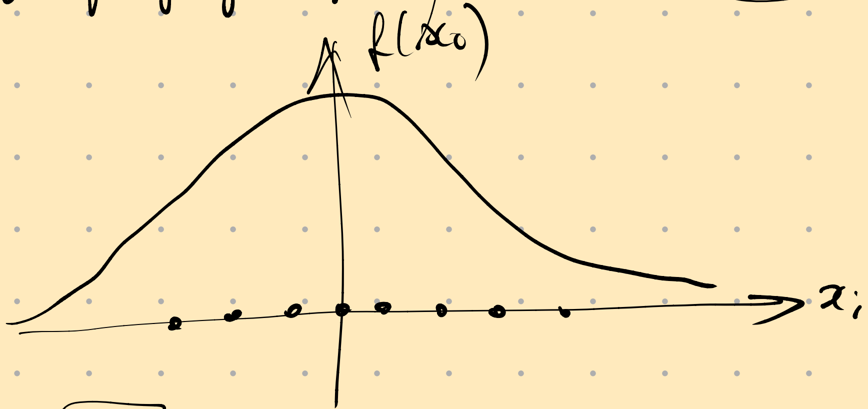
$$\hat{\mu} = \frac{\sum x_i}{n}$$

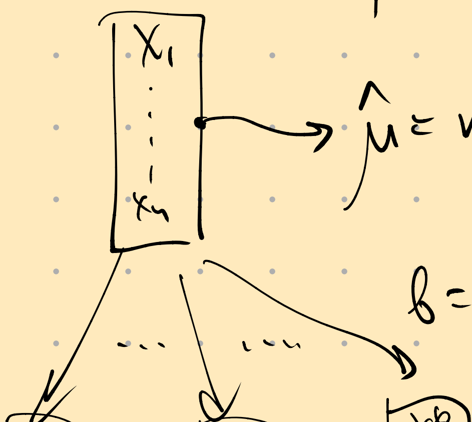
$$\hat{\mu} = \begin{cases} \frac{\sum x_i}{n}, & \text{если } \frac{\sum x_i}{n} \geq 0 \\ 0, & \text{если } \frac{\sum x_i}{n} < 0 \end{cases}$$

а) $\hat{\mu} = \max\{0, \frac{\sum x_i}{n}\}.$

б) что происходит с наивным бутстрэпом? $n >> 0$

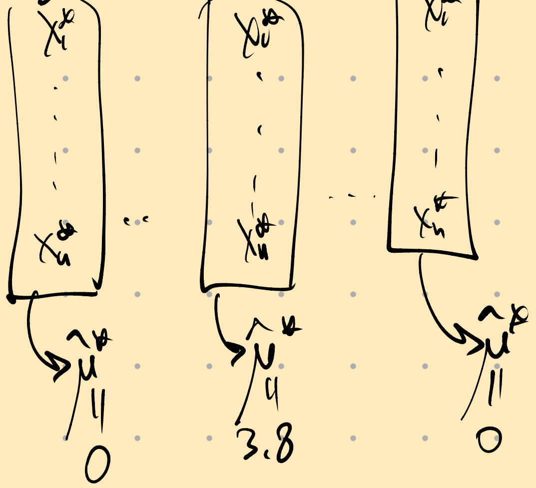
что произойдёт при $\mu = 0$?





$\hat{\mu} = \max\{\bar{X}, 0\}$

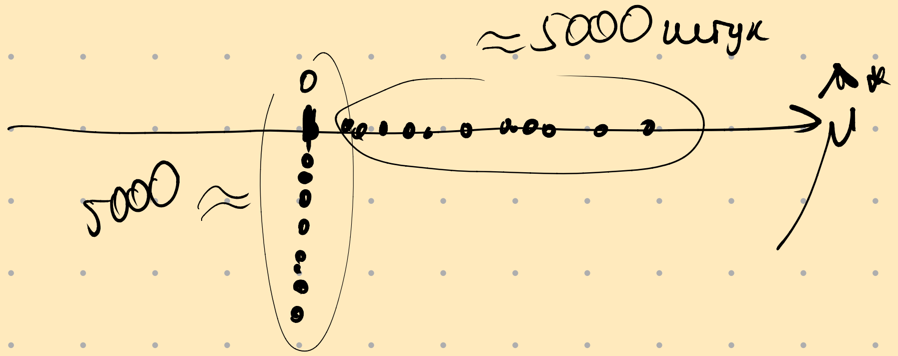
$B = 10000$



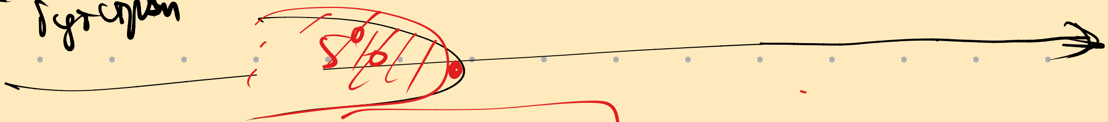
закон распр. для $X_1^* \ldots X_n^*$ симм-н отн $\bar{X}$.

(n - велико)

$B = 10000$



наивн. бутстрап



почти гар-но $\boxed{\hat\mu_L^{boot} = 0}$

ожидаю $P(\mu \ge \hat\mu_L^{boot}) = 95\%$

нам-но одност. 95% CI

фактич. $P(\mu \ge \hat\mu_L^{boot}) = 100\%$