Лекция 15

• доверительный интервал для медианы

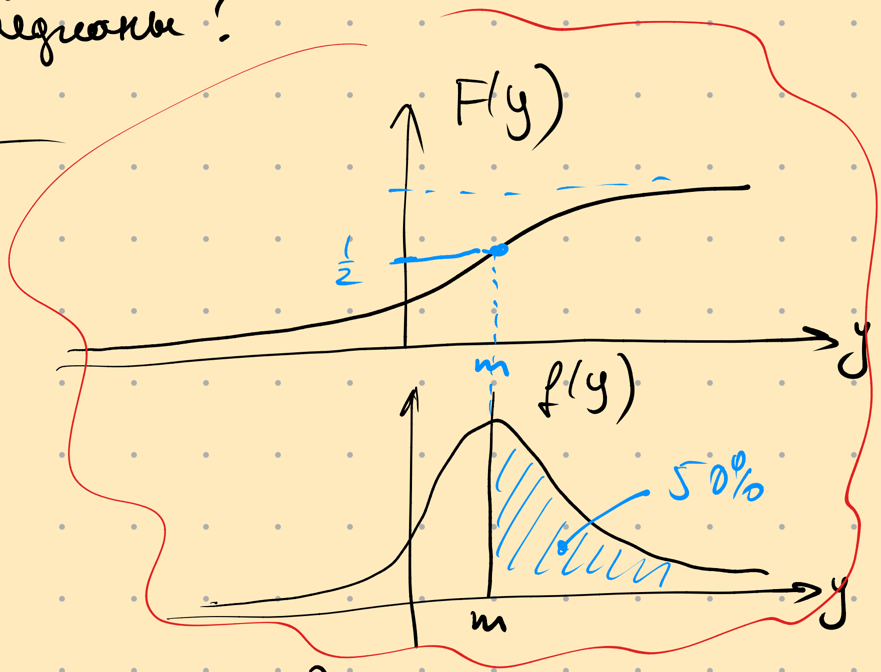
Задача.

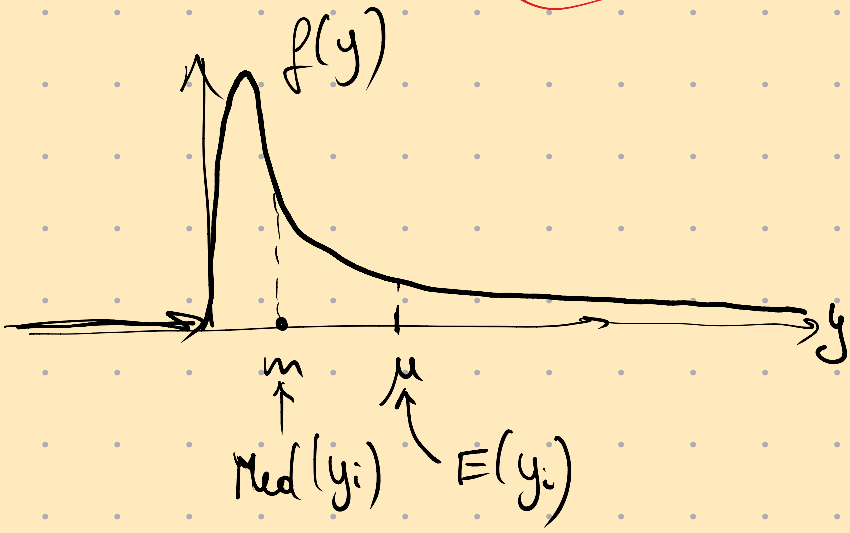
$y_1, y_2 \dots y_n$ ~ независимы, любое непрерывное расп.

!! а) точечная оценка для $Med(y_i) = m$?

б) довер. инт для медианы? 90%?

набл: $y_1, y_2, y_3 \dots y_n$





$$\hat{\mu} = \frac{y_1 + y_2 + \dots + y_n}{n}$$

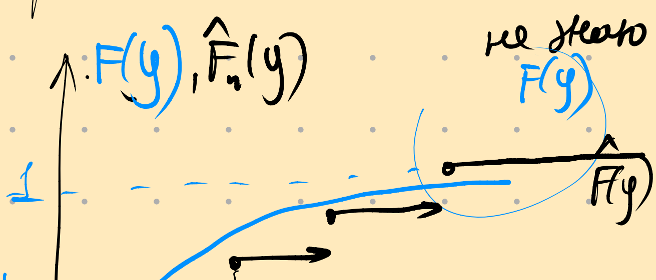
$$\operatorname{plim} \hat{\mu} = \operatorname{plim} \bar{y} = E(y_i)$$

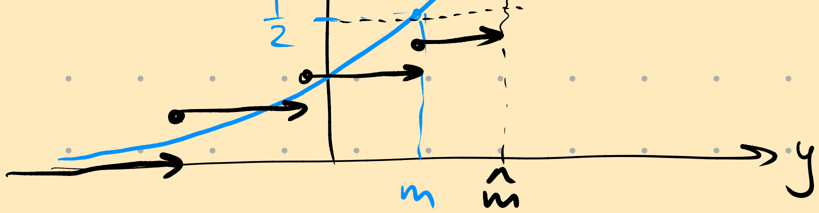
$\hat{m}$?

хочу

$$\operatorname{plim} \hat{m} = m = Med(y_i)$$

$$F(y) = P(\overset{\text{св.}}{y_i} \le \overset{\text{const}}{y})$$





def Выборочная ф-ция распределения

$$\hat F_n(y) = \hat P(y_i \le y) = \frac{\text{кол-во набл-ий} \le y}{n}$$

↖ общее кол-во наблюдений.

$\boxed{\text{Утв}}$ $\quad \operatorname{plim}_{n\to\infty} \hat F_n(y) = F(y)$

$\boxed{\text{док-во !!}}$

$y = 42$!! $\qquad F(42) = P(y_i \le 42)$

$$\hat F_n(42) = \frac{\text{кол-во набл} \le 42}{n} =$$

$$= \frac{I_1 + I_2 + \ldots + I_n}{n}$$

$$I_1 = I(y_1 \le 42)$$

ЗБЧ

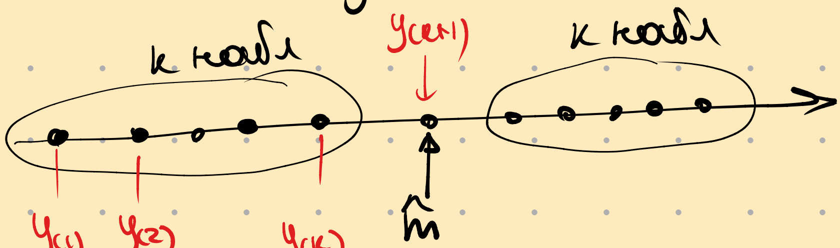
$$\operatorname{plim} \hat F_n(42) = \operatorname{plim} \frac{I_1 + \ldots + I_n}{n} = E(I_1) =$$

$$= P(I_1 = 1) = P(y_1 \le 42)$$

на практике:

→ отсортировать $y_1 \ldots y_n$ по возраст-ию

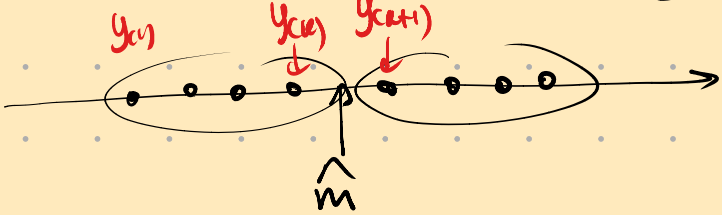
$$y_{(1)} \le y_{(2)} \le \ldots \le y_{(n)}$$



n - нечётное

$n = 2k+1$

$$\hat m = \begin{cases} y_{(k+1)}, & \text{если } n = 2k+1 \\ \dfrac{y_{(k)} + y_{(k+1)}}{2}, & \text{если } n = 2k \end{cases}$$



n - чётное

$n = 2k$

Упр. $n=11$ $y_1 \dots y_{11}$ незав, непр-но распр-ные один-во.

| 2 | 7 | 17 | 33 | 11 | 28 | 13 | 8 | 15 | 6 | 19 |

точки: $2, 6, \underset{y_{(3)}}{7}, 8, 11, 13, 15, 17, \underset{y_{(9)}}{19}, 28, 33$

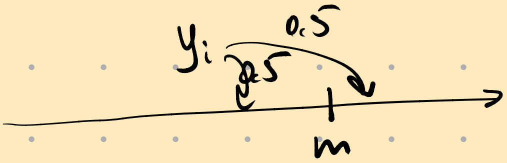
$\hat{m} = 13$

честно

90%

$$\boxed{P(y_{(3)} \le m \le y_{(9)})} = ?$$

$$X_i = \begin{cases} 1, & \text{если } y_i > m \\ 0, & \text{если } y_i < m \end{cases}$$



$y_1 \dots y_n$ ← видим

$x_1 \dots x_n$ ← не видим

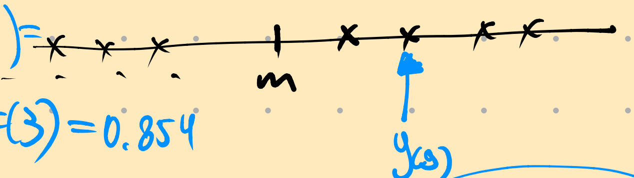
$X_i \sim$ незав $P(X_i=1) = P(X_i=0) = \frac{1}{2}$

кол-во наблюд-ий правее m → $S = X_1 + \dots + X_{11}$ $S \sim Bin(n=11, p=\frac{1}{2})$

$$P(y_{(3)} \le m \le y_{(9)}) = P(S \in ?) = P(3 \le S \le 8) = P(S \le 8) - P(S \le 2) = (cdf) = 0.93 = 0{,}93457$$

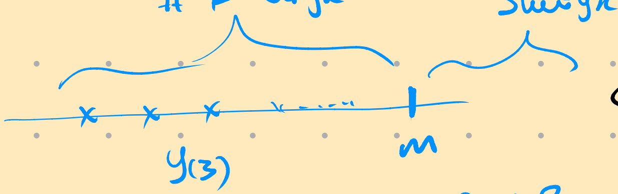
$m \le y_{(9)} \iff S \ge 3$

$$P(y_{(3)} \le m \le y_{(8)}) = P(S \in [4;8]) = P(S \le 8) - P(S \le 3) = F(8) - F(3) = 0.854$$



$y_{(3)} \le m \iff$ левее медианы попало хотя бы 3 игрека $\iff$ правее медианы попало не более 8 игреков.

$S \le 8$



$11 - S \ge 3$ $S \le 8$

• Разные интервалы.

доверительные интервалы
- confidence interval
- predictive interval
- tolerance interval.

с мат. точки зрения разницы почти нет

$$P(\underline{Y} \in [T_L; T_H]) = 0{,}95$$

неизв. пар-р. / будущее наблюдение.

Пример.

$X_i \sim$ нерав $U[0; a]$

$\boxed{\text{наблюдаю } X_1}$

• Хочу: довер. инт для a 95%
conf. interval a.

• Хочу: довер. инт / предсказ / предикт. X_2
predictive interval

• Хочу: дов. инт.: накрыл 80% буд. наблюдений. 95%
tolerance interval

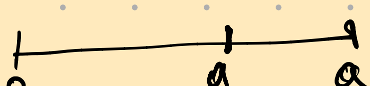
а). $P(a \in [T_L; T_H]) = 0{,}95$ CI for a

б). $P(X_2 \in [T_L; T_H]) = 0{,}95$ PI for X_2

в). $P(80\% \text{ от } X_2, X_3, \ldots\ldots \in [T_L; T_H]) = 0{,}95$ TI for 80% fut. values

$[T_L; T_H] = [0; k\cdot X_1]$

а) $P(a \in [0; kX_1]) =$
$= P(a \le k\cdot X_1) = P(X_1 \ge a/k) =$



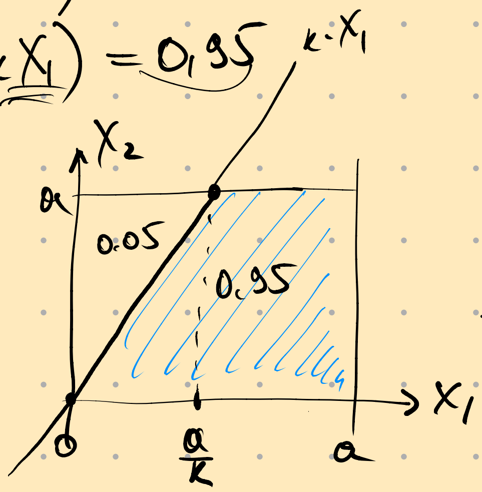
$$= \frac{\left(a-\frac{a}{k}\right)}{a} = 1-\frac{1}{k} = 0{,}95$$

$$\frac{1}{k} = 0{,}05 \qquad k=20$$

$\boxed{95\% \text{ инт-л для } a \quad [0;\, 20X_1]}$

б) $P(X_2 \in [0;\, kX_1]) = 0{,}95$

$P(X_2 \le kX_1) = 0{,}95$



$P(X_2 > kX_1) = 0{,}05$

$$\frac{\frac{1}{2}\cdot a \cdot \frac{a}{k}}{a^2} = \frac{S}{S} = 0{,}05$$

$$\frac{1}{2k} = 0{,}05$$

$$k = \frac{1}{2}\cdot\frac{1}{0{,}05} = \frac{1}{2}\cdot 20 = 10$$

PI for X_2 95%

$P(X_2 \in [0;\, 10X_1]) = 0{,}95$

Задача (интерпр)

незав. набл. $y_1 -$, $y_2 -$, …, $y_n -$ цены 2-х кв.: $P_{(\text{усл.})}$.

Что хотим накрыть? 95%

① цену след-ей квартиры (подх-ей под условия)

② среднюю по москве цену такой квартиры

→ ① $P(y_{n+1} \in [T_L(y_1 \dots y_n);\, T_H(y_1 \dots y_n)]) = 0{,}95$

2 $P\left(\underline{E(y_i)} \in [\, l_L(y_1, \dots, y_n) ;\ u(y_1, \dots, y_n) \,] \right) = 0.95$