4.1 | Дядя Фёдор

Классический подход:

θ - неизвестный параметр (константа)

$X_1, \dots, X_n$ - выборка

$\hat{\theta} = f(X_1, \dots, X_n)$ - оценка (сл. вел.)

Разные методы:

$\hat{\theta}^{MM}$ $\hat{\theta}^{ML}$ - точечные оценки

Упражнение

Гипотезово:

Маша	140
Паша	150
Саша	200
Даша	190
вся ген. сов.	рост

μ - ср. рост

Матроскин $\longrightarrow$ Маша, Паша $\hat{\mu}^{M}$

Шарик $\longrightarrow$ Саша, Даша $\hat{\mu}^{Ш}$

Фёдор $\longrightarrow$ Маша, Саша $\hat{\mu}^{Ф}$

едут через Гипотезово в Простоквашино

а) Настоящее μ - ? ; оценки парней

$$\mu = \frac{140 + 150 + 200 + 190}{4} = 170$$

$\hat{\mu}^{Ф} = 170$

$\hat{\mu}^{Ш} = 195$

$\hat{\mu}^{M} = 145$

б) Посорятся ли?

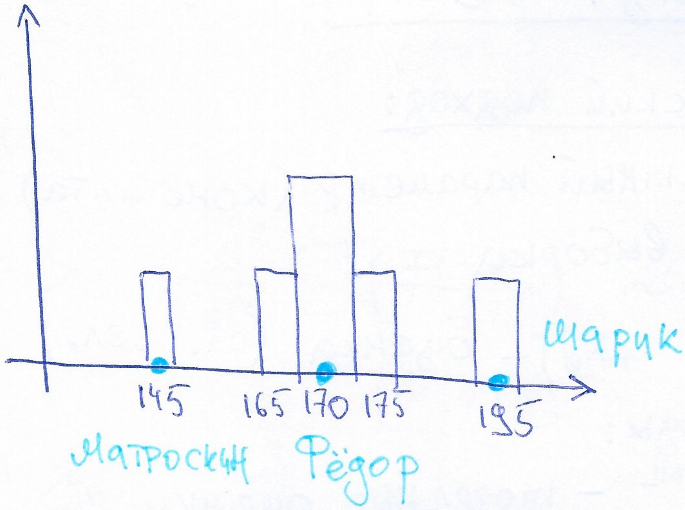
Да

в) Печкину надоели ссоры. Поехал в гипотезово. Сколько $\exists$ вариантов выборок - ?

$$C_4^2 = \frac{4!}{2!\,2!} = 6.$$

г) Найти распределение $\bar{X}$ по 2 наблюдениям.

Маша	Паша	145
Маша	Саша	170
Маша	Даша	165
Паша	Саша	175
Паша	Даша	170
Саша	Даша	195

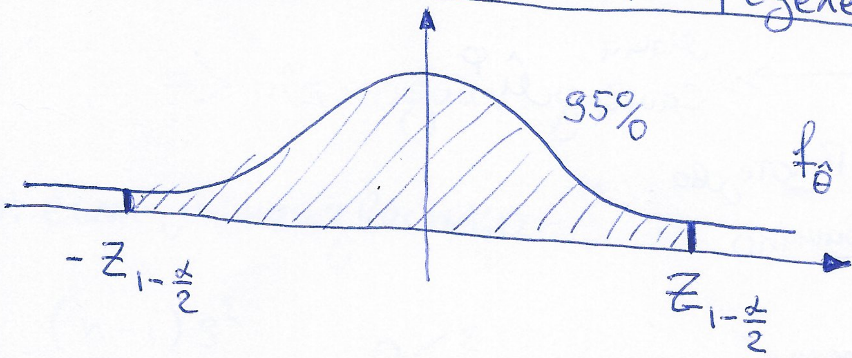


Фёдору просто повезло.
Шарику и Матроскину нет!!

А вдруг мы тоже строили оценку и нам не повезло?

$\Rightarrow$ доверительные интервалы.

$\hat{\theta}$ – сл. вел. с каким-то распределением.



$$\hat{\theta} = \bar{X} \overset{asy}{\sim} N(\theta, Var(\hat{\theta}))$$

вычитаем мат. ож. – центрирование.

делим на se – нормирование

$$\frac{\hat{\theta} - \theta}{se(\hat{\theta})} \overset{asy}{\sim} N(0, 1)$$

$$z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$$

9) Печкин строит дов. интервал.

$$\mathbb{P}\left(-1.96 \leq \frac{\bar{X} - \mu}{se(\bar{X})} \leq 1.96\right) = 0.95$$

Если сделаем это, все три интервала накроют 170.

Для Фёдора $se(\bar{X}) = 29.7$

Прошла неделя. Пьют чай. Вопрос у Фёдора: А может ли быть такое, что $\mu = 160$?

- $\bar{X}$ – сл. вел.
- $\bar{X} - 160$ – сл. вел.
- Если гипотеза H_0 верна и $\mu = 160$, тогда $\bar{X} - 160$ концентрируется около нуля.
- Строим дов. инт. для $\bar{X} - 160$. Если 0 в нём $\Rightarrow$ гипотеза не отвергается.

$$\bar{X} \overset{asy}{\sim} N(\mu, Var(\bar{X}))$$

$H_0: \mu = 160$

$H_A: \mu \neq 160$.

Если наши домыслы правда, тогда $\bar{X} \overset{asy}{\sim} N(160, (29.7)^2)$

$$Z = \frac{\bar{X} - 160}{29.7} \sim N(0, 1)$$

$$Z_{набл.} = \frac{\overset{\hat{\mu}^{Ф.}}{170} - 160}{29.7} = ?$$

попала $\Rightarrow$ не отвергаем.

$[-1.96; 1.96]$.

~~Схема:~~

~~$\hat{\theta}$ оценка точечная $\xrightarrow{?}$ распределение $\xrightarrow{\text{если поняли}}$ доверительный интервал~~

Схема Матстата (что хочет статистик?)

$\hat{\theta}$ $\longrightarrow$ $f_{\hat{\theta}}$ $\longrightarrow$ (дов. инт.) понимает насколько оценка точная | может отвечать на вопросы (гипотезы)

точечная оценка — распределение оценки

Как получить:
- метод моментов
- метод макс. правд.

Разные теоремы:
- ЦПТ, дельта-метод — асимптотика
- $\boxed{\text{Т. Фишера, Р-е. Стьюдента.}}$ — точные
- + другие

У оценок должны быть хорошие свойства
$\Downarrow$
выбор метода для оценки

- Несмещенность
  $\mathbb{E}(\hat{\theta}) = \theta$. $\quad \underbrace{bias(\hat{\theta})}_{\text{смещение}} = \mathbb{E}(\hat{\theta}) - \theta$
- Состоятельность.
  $\hat{\theta} \xrightarrow{P} \theta$
- Эффективность.
  $\forall \tilde{\theta}: \; bias(\tilde{\theta}) = bias(\hat{\theta})$ $\quad Var(\hat{\theta}) \leq Var(\tilde{\theta})$
  чем меньше дисперсия, тем уже доверительный интервал и точнее вывод
- Знать распределение $\hat{\theta}$

Упражнение

а) $X_1, \ldots, X_n \sim iid\, N(\mu, 1)$

$\hat{\mu}$ — ?

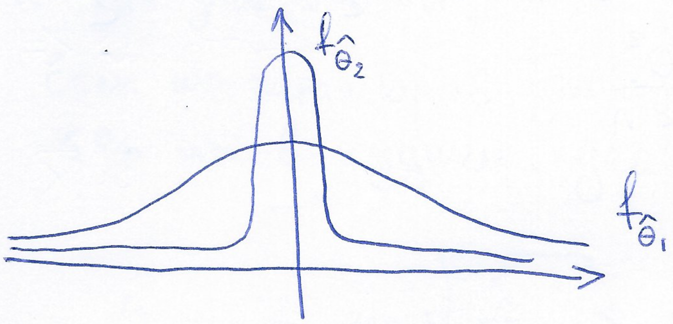
	несмещ.	смещ.
сост.	$\overline{X}$	$\overline{X} + \frac{1}{n}$
несост.	X_1	$arctg(\overline{X})$

$X_1, \dots, X_n \sim iid\ U[0;a]$

$\hat{a}$ – ?

	несмещ.	смещ.
сост.	$2\overline{X}$	x_{max}
несост.	$2X_1$	$x_1^{x_2} - 8$

Упражнение



Какая эффективнее?

$\hat{\theta}_1$ или $\hat{\theta}_2$

Обычно, чтобы сравнивать оценки между собой используют $MSE(\hat{\theta}) = \mathbb{E}((\hat{\theta} - \theta)^2)$.

$$MSE(\hat{\theta}) = \mathbb{E}((\hat{\theta} - \theta)^2) = \mathbb{E}((\hat{\theta} - \tilde{\theta} + \tilde{\theta} - \theta)^2) =$$

$$\tilde{\theta} = \mathbb{E}(\hat{\theta})$$

$$= \underbrace{\underbrace{\mathbb{E}(\hat{\theta} - \tilde{\theta})^2}_{\mathbb{E}(\hat{\theta} - \mathbb{E}(\hat{\theta}))^2}}_{Var(\hat{\theta})} + 2(\tilde{\theta} - \theta) \cdot \underbrace{\underbrace{\mathbb{E}(\hat{\theta} - \tilde{\theta})}_{\mathbb{E}(\hat{\theta}) - \mathbb{E}(\hat{\theta})}}_{0} + \underbrace{\underbrace{\mathbb{E}(\tilde{\theta} - \theta)^2}_{(\mathbb{E}(\hat{\theta}) - \theta)^2}}_{[bias(\hat{\theta})]^2}$$

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + [bias(\hat{\theta})]^2$$

Упражнение

$X_1, \dots, X_n \sim iid\ U[0;\theta]$.

$\hat{\theta} = \bar{X}$

$bias(\hat{\theta})\quad Var(\hat{\theta})\quad MSE(\hat{\theta}) - ?$

$\mathbb{E}(X) = \frac{\theta}{2} \qquad Var(X) = \frac{\theta^2}{12}.$

$\mathbb{E}(\hat{\theta}) = \mathbb{E}(\bar{X}) = \frac{1}{n}\sum_i \mathbb{E}(X_i) = \frac{\theta}{2}. \Rightarrow bias = \frac{\theta}{2} - \theta = -\frac{\theta}{2}$

$Var(\hat{\theta}) = Var(\bar{X}) = \frac{1}{n^2}\sum_i Var(X_i) = \frac{\theta^2}{12n}$

$MSE(\hat{\theta}) = \frac{\theta^2}{12n} + \frac{\theta^2}{4}.$

$\boxed{\text{Блокнот про св-ва оценок}}$

4.3 Доверительные интервалы на основе ЦПТ

Как понять какое у статистики распределение?
Спросить союзников!

Упражнение

Ульяна Игра Престолов.

X – число серий, которые она посмотрела за день.

$X_1, \dots, X_n \sim Poiss(\lambda)$ λ – интенсивность.

а) $\hat{\lambda}^{MM}$

$$\underbrace{\mathbb{E}(X)}_{\lambda} = \bar{X} \Rightarrow \hat{\lambda} = \bar{X}$$

$Var(X) = \lambda$

б) Как распределена $\hat{\lambda}$ - ?

Не знаем, но хотели бы $\Rightarrow$ заменим на $\hat{\lambda}$.

$$\hat{\lambda} = \bar{X} = \frac{X_1 + \dots + X_n}{n} \overset{asy}{\sim} N\left(\lambda, \frac{\lambda}{n}\right)$$

(Т. Слуцкого о наследовании сходимости)

$$\frac{\hat\lambda-\lambda}{\sqrt{\frac{\hat\lambda}{n}}} \overset{asy}{\sim} N(0,1)$$

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\hat\lambda-\lambda}{\sqrt{\frac{\hat\lambda}{n}}} \leq z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

α - ур. значимости.

Если 100 парашютов упаковали на уровне значимости 5% $\Rightarrow$ в худшем случае умрут пятеро.

$$\hat\lambda - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat\lambda}{n}} \leq \lambda \leq \hat\lambda + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat\lambda}{n}}$$

$\boxed{\text{На компе}}$

в) У Ульяны биполярочка - Таня

девушки независимые λ_1 λ_2

$\hat\lambda_1 = \bar{x}$ $\hat\lambda_2 = \hat{y}$

дов. инт. для $\hat\lambda_1 - \hat\lambda_2$

Если 0 внутри $\Rightarrow$ одинаково

$$\hat\lambda_1 \overset{asy}{\sim} N\left(\lambda_1; \frac{\lambda_1}{n_1}\right)$$

$$\hat\lambda_2 \overset{asy}{\sim} N\left(\lambda_2; \frac{\lambda_2}{n_2}\right)$$

$$\hat\lambda_1 - \hat\lambda_2 \sim N\left(\lambda_1-\lambda_2; \frac{\lambda_1}{n_1}+\frac{\lambda_2}{n_2}\right)$$

$$\mathbb{E}(\hat\lambda_1-\hat\lambda_2) = \lambda_1-\lambda_2$$

$$Var(\hat\lambda_1-\hat\lambda_2) = Var(\hat\lambda_1)+Var(\hat\lambda_2) = \frac{\lambda_1}{n_1}+\frac{\lambda_2}{n_2}.$$

$$\hat\lambda_1 - \hat\lambda_2 \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat\lambda_1}{n_1}+\frac{\hat\lambda_2}{n_2}}$$

$\boxed{\text{На компе}}$

В самом начале мы заменили $N(\lambda, \frac{\lambda}{n})$ на $N(\lambda, \frac{\hat\lambda}{n})$.

$$\underbrace{\underbrace{\frac{\hat\lambda-\lambda}{\sqrt{\frac{\lambda}{n}}}}_{N(0,1)} \cdot \underbrace{\frac{\sqrt{\frac{\lambda}{n}}}{\sqrt{\frac{\hat\lambda}{n}}}}_{1}}_{N(0,1)} = \underbrace{\underbrace{\frac{\hat\lambda-\lambda}{\sqrt{\frac{\hat\lambda}{n}}}}_{?} \cdot \underbrace{\frac{\sqrt{\frac{\hat\lambda}{n}}}{\sqrt{\frac{\lambda}{n}}}}_{\xrightarrow{P} 1 \text{ т.к. } \hat\lambda \xrightarrow{P} \lambda}}_{N(0,1)}$$

$$\Rightarrow ? \sim N(0,1)$$

Мы строго обосновали замену λ на $\hat\lambda$

А м.б. можно ещё как-то ~~ценить~~ заменить?

$\boxed{4.4}$ Другие дов. интервалы

Почему мы заменили $N(\lambda, \frac{\lambda}{n})$ на $N(\lambda, \frac{\hat\lambda}{n})$.

$$\underbrace{\underbrace{\frac{\hat\lambda-\lambda}{\sqrt{\frac{\lambda}{n}}}}_{N(0,1)} \cdot \underbrace{\frac{\sqrt{\frac{\hat\lambda}{n}}}{\sqrt{\frac{\hat\lambda}{n}}}}_{1}}_{N(0,1)} = \underbrace{\underbrace{\frac{\hat\lambda-\lambda}{\sqrt{\frac{\hat\lambda}{n}}}}_{?} \cdot \underbrace{\frac{\sqrt{\frac{\hat\lambda}{n}}}{\sqrt{\frac{\lambda}{n}}}}_{\xrightarrow{P} 1 \text{ т.к. } \hat\lambda \xrightarrow{P} \lambda}}_{N(0,1)}$$

$$\Rightarrow ? \sim N(0,1).$$

1.4 Дельта-метод (обобщение ЦПТ)

Обычно метод моментов даёт оценки, похожие на $\frac{\overline{x^2} - \overline{x}}{\overline{\ln x}}$; т.е. это какая-то сложная ф. от среднего.

Среднее - клёвое. Для него есть ЦПТ. Значит для таких штук тоже что-то должно быть!

Т. (дельта-метод)

Если $X_1, \dots, X_n$ нез. и одинаково распределены, причём $\mathbb{E}(X_i) = \mu$, $Var(X_i) = \sigma^2$, а $g(t)$ - диф. ф. тогда

$$g(\bar{X}) \overset{asy}{\sim} \mathcal{N}\left(g(\mu); \frac{\sigma^2}{n} \cdot (g'(\mu))^2\right)$$

Упражнение

$X_1, \dots, X_{100} \sim U[2; 8]$ и нез.

$\mathbb{E}(X_i) = 5$

$Var(X_i) = \frac{(8-2)^2}{12} = 3$

а) $\bar{X}_{100} \sim ?$

б) $\frac{1}{\bar{X}_{100}} \sim ?$

$$\bar{X}_{100} \overset{asy}{\sim} \mathcal{N}\left(5; \frac{3}{100}\right)$$

$g(t) = \frac{1}{t}$

$g'(t) = -\frac{1}{t^2}$

$$\frac{1}{\bar{X}_{100}} \overset{asy}{\sim} \mathcal{N}\left(\frac{1}{5}; \frac{3}{100} \cdot \left(\frac{1}{25}\right)^2\right)$$

Почему?

можно разложить $\frac{1}{\bar{x}}$ в ряд Тэйлора в окрестности $\mathbb{E}(\bar{x})$.

$$g(\bar{x}) = \frac{1}{\bar{x}} = \frac{1}{5} - \frac{1}{25}(\bar{x}-5) + o(\bar{x}).$$

$$g(t) = g(t_0) + g'(t_0)(t-t_0) + o(t).$$

$$\mathbb{E}\left(\frac{1}{\bar{x}}\right) \approx \frac{1}{5}$$

$$\mathrm{Var}\left(\frac{1}{\bar{x}}\right) \approx \left(\frac{1}{25}\right)^2 \cdot \mathrm{Var}(\bar{x}-5) = \left(\frac{1}{25}\right)^2 \cdot \overset{3/100}{\overline{\mathrm{Var}(\bar{x})}}$$

2). Ульяна

$$\hat{\lambda}^2 = (\bar{x})^2$$

$$\mathbb{E}(X) = \mathrm{Var}(X) = \lambda$$

$$g(t) = t^2 \quad g(\mu) = \lambda^2$$

$$g'(t) = 2t \quad g'(\mu) = 2\lambda$$

$$\hat{\lambda}^2 \overset{asy}{\sim} N\left(\lambda^2; \frac{\lambda}{n} \cdot (2\lambda)^2\right).$$

$$\bar{x} \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{4\bar{x}^3}{n}}$$

.5 Точные доверительные интервалы

Какое у $\hat{\theta}$ распределение?

- ЦПТ
- Δ-метод

| Асимптотически нормальное для средних.

Иногда можно понять точное р.-е. для $\hat{\theta}$.

① $X \sim N(\mu, \sigma^2)$

$\hat{\mu} = \overline{X}$ $\qquad \hat{\mu} \overset{asy}{\sim} N\left(\mu, \frac{\hat{\sigma}^2}{n}\right)$.

Если σ^2 известно, тогда $\hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

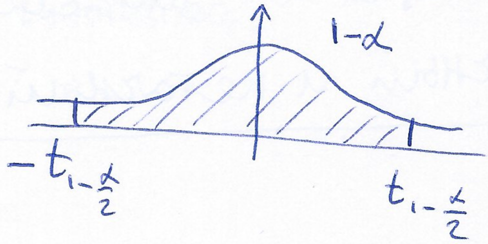
А можно ли узнать точный интервал, если σ^2 неизвестно?

$$\underbrace{\frac{\hat{\mu}-\mu}{\frac{\sigma}{\sqrt{n}}} : \sqrt{\frac{1}{n-1} \cdot \frac{(n-1)s^2}{\sigma^2}}}_{\frac{\hat{\mu}-\mu}{s/\sqrt{n}}} \sim \frac{N(0,1)}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}} = t(n-1)$$

Т. Фишера

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\Rightarrow \overline{x} \pm t_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$

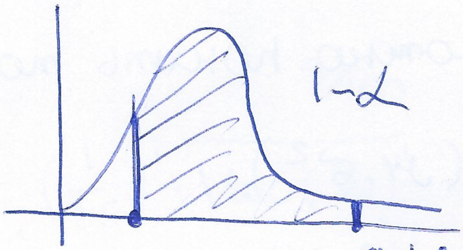


② $\hat{\sigma}^2 = S^2$

Пусть мы знаем μ, тогда:

$$S^2 = \frac{1}{n}\sum_i \underbrace{(x_i - \mu)^2}_{[N(0,\sigma^2)]^2} \Rightarrow n\cdot\frac{S^2}{\sigma^2} \sum_i \underbrace{\left(\frac{x_i-\mu}{\sigma}\right)^2}_{[N(0,1)]^2} \sim \chi^2_n$$

1 тогда χ^2_n



qchisq(0.025) qchisq(0.975)

$$\chi^2_n\left(\frac{\alpha}{2}\right) \le \frac{nS^2}{\sigma^2} \le \chi^2_n\left(1-\frac{\alpha}{2}\right)$$

$\Rightarrow$ Находим для σ^2.

А если μ неизвестно — ?

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \text{ по Т. Фишера}$$

$$\Rightarrow \chi^2_{n-1}\left(\frac{\alpha}{2}\right) \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{n-1}\left(1-\frac{\alpha}{2}\right)$$

Время? $\Rightarrow$ сгенерить в R ген. сов. из N( )
2 интервала 100, 200, их длины.
точный и неточный