Esercitazione 6 - Point Estimate and Two-sided confidence intervals

Exercise 1

$X_1, \dots, X_n$ (iid)

$\mu = 5.7 \quad \sigma = 4,1$

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

a) $\mathbb{E}[X_n]$? $\mathrm{Var}(\bar{X}_n)$? $(n = 4,\ n = 100)$

$$\mathbb{E}[\bar{X}_n] \underset{\text{LUN}}{=} \frac{1}{n}\sum_{i=1}^{n} \underbrace{\mathbb{E}[X]}_{\mu} = \mu$$

$$\mathrm{Var}[\bar{X}_n] = \frac{1}{n}\sigma^2 \longrightarrow$$

$\rightarrow n = 4 \Rightarrow 4,2025$

$\rightarrow 0,1681$; $n = 100$

↳ more data, smaller variance.

$$\mathbb{E}[\bar{X}_n] = 5,7 \quad \forall n$$

b) $X_1, \dots, X_4 \overset{iid}{\sim} \mathcal{N}(5.7, 4.1^2)$

$$\mathbb{P}[\bar{X}_n < 6.5] = ?$$

$$\bar{X}_4 \sim \mathcal{N}(5.7,\ 4,2025)$$

$$\mathbb{P}\left[\frac{\bar{X}_4 - 5,7}{\sqrt{4,2025}} \le \frac{6,5 - 5,7}{\sqrt{4,2025}}\right) = \phi(0,39) = 0,65173$$

c) $\mathbb{P}(\bar{X}_{100} \leq 6{,}5) = ?$

Since we don't know gaussianity we have to apply CLT

$$\left.\begin{array}{l} n = 100 \geq 50 \\ X_1, \dots, X_{100} \text{ iid} \end{array}\right\} \Rightarrow \text{we can apply CLT} \Rightarrow \bar{X}_n \overset{\text{approx}}{\sim} N(5.7, 0{,}1681)$$

$$\mathbb{P}(\bar{X}_{100} \leq 6{,}5) \simeq \phi(1.95) = 0{,}97441$$

Exercise 2 → Stat. Point wise and Interval Estimator

We will always assume:

$X_1, \dots, X_n \overset{\text{iid}}{\sim}$

$X \sim f$

$\hat{\Theta} \longrightarrow \hat{\theta}$

$\downarrow$ estimator (r.v) ↳ function of the sample

$\downarrow$ estimate (number) ↳ computable only with data

$$f(x) = \frac{1}{3\theta} e^{\frac{-x}{3\theta}} \mathbb{I}_{(0,\infty)}(x) \rightarrow \sim Exp\left(\frac{1}{3\theta}\right)$$

$\theta > 0$, unknown parameter

a) $X \sim f \Rightarrow X \sim Exp\left(\frac{1}{3\theta}\right) \rightarrow \lambda = 1/3\theta$

$\rightarrow$ Compute μ, σ^2 $\rightarrow \mu = 1/\lambda = 3\theta$

$\hookrightarrow \sigma^2 = (3\theta)^2 = 9\theta^2$

$$\mu = \int_{\mathbb{R}} x f(x)\, dx$$

b) $\sum_{j=1}^{10} X_j \rightarrow$ Statistics $=300$

Estimate μ, σ^2

$$T = \sum_{j=1}^{10} X_j$$

$$\bar{X}_n = \frac{1}{n}\sum_{j=1}^{10} X_j = \frac{T}{10} \rightarrow \text{Estimator for } \mu$$

$$\bar{X}_n = \frac{300}{10} = 30 \rightarrow \text{Estimate for } \mu$$

$$\frac{1}{n-1}\sum_{j=1}^{n}(X_j - X_n)^2 = \frac{1}{n-1}\left(\sum_{j=1}^{n} X_j^2 - n\bar{X}_n^2\right)$$

$\rightarrow$ Can't be used since we don't have stastic.

$\rightarrow$ Not the right way.

$$\sigma^2 = 9\theta^2 = \mu^2$$

Plug-in rule

since $\sigma^2 = \mu^2 \Rightarrow$ Estimator for $\sigma^2 = (\text{Estimator for } \mu)^2$

$$S^2 = X_n^2 = \frac{T^2}{100} \rightarrow \text{Estimator for } \sigma^2$$

$s^2 = 30^2 = 900 \rightarrow$ Estimate of variance.

c) Construct $\hat{\Theta}$ for θ; computer bias, mean square error, standard error

$$\mu = E[X] = 3\theta \Rightarrow \theta = \frac{\mu}{3} \rightarrow \text{Plug-in rule}$$

$$\hat{\Theta} = \frac{\bar{X}_n}{3} \rightarrow \text{Estimator for } \theta$$

$$b_\theta(\hat{\Theta}) = E[\hat{\Theta}] - \theta = E\left[\frac{\bar{X}_n}{3}\right] - \theta = \frac{\mu}{3} - \theta = \theta - \theta = 0$$

$\Rightarrow$ Estimator is

unbiased.

$$MSE_\theta(\hat{\Theta}) = E\left[(\hat{\Theta}-\theta)^2\right] = \cancel{b_\theta^2(\hat{\Theta})}^{\to 0} + Var(\hat{\Theta})$$

$$MSE(\hat{\Theta}) = Var(\hat{\Theta}) = \frac{1}{9}\frac{\sigma^2}{n} = \frac{1}{9}\frac{\mu^2}{n} = \frac{\theta^2}{n}$$

$$\underbrace{se}_{\text{Standard Error}} = \sqrt{Var(\hat{\Theta})} = \theta/\sqrt{n}$$

d) Estimate Standard error using $\sum_{j=1}^{10} X_j = 300$

Plug-in rule

$$\hat{se}(\hat{\theta}) = \frac{\hat{\theta}}{\sqrt{n}} = \frac{\bar{X}_n}{3\sqrt{n}}$$ → Estimator for standard error

An estimate is $\frac{30}{3\sqrt{10}} = \frac{10}{\sqrt{10}} = \sqrt{10}$

Exercise 6.3

$X_1, \dots, X_{20} \overset{iid}{\sim}$, μ, σ^2 unknown

$$\sum_{j=1}^{20} x_j = 600 \ ; \ \sum_{j=1}^{20} x_j^2 = 18142$$

a) $\bar{X}_{20} = \frac{1}{20}\sum_{j=1}^{20} X_j$ → Estimator for μ (unbiased)

Sample Variance → $S^2 = \frac{1}{19}\sum_{j=1}^{20}(X_j - \bar{X}_{20})^2 = \frac{1}{19}\left(\sum_{j=1}^{20} X_j^2 - 20\sum \bar{X}_{20}^2\right)$

↳ Unbiased estimator of variance → Estimator for σ^2 (unbiased)

Assume $\mu = \sqrt{\theta}$, $\theta > 0$

b) $\hat{\Theta}$? $\hat{\theta}$?

$\mu = \sqrt{\theta} \Rightarrow \theta = \mu^2$; By plug-in rule $\hat{\Theta} = \bar{X}_n^2 = \dfrac{\left(\sum_{j=1}^{20} X_j\right)^2}{400}$

↳ Estimator for θ, when $\theta = \mu^2$

$\hat{\theta} = \dfrac{600^2}{400} = 900$ (Estimate for θ)

c) $b_\theta(\hat{\Theta})\overset{?}{=} E[\hat{\Theta}] - \theta = \underbrace{E[\bar{X}_n^2]}_{Var(\bar{X}_n) + E^2(\bar{X}_n) = \frac{\sigma^2}{20} + \mu^2} - \theta$

$\overset{*}{=} \dfrac{\sigma^2}{20} + \theta - \theta = \dfrac{\sigma^2}{20} \neq 0$ (biased)

<u>Exercise 6,4</u> (Interval Estimators)

X, weight empty box (kg)

Y, weight full box (kg)

$X_1, X_2, X_3 \overset{iid}{\sim} N(\mu_X, 0{,}1)$

$Y_1, \dots, Y_5 \overset{iid}{\sim} N(\mu_Y, 0{,}1)$

$X \perp Y$

$\delta := \mu_Y - \mu_X$

$\hat{\Delta}_1 = \bar{Y} - \bar{X}$

$\hat{\Delta}_2 = Y_1 + Y_2 + Y_3 - (Y_4 + Y_5) + X_1 - (X_2 + X_3)$

a) $b_\delta(\hat{\Delta}_1)$?, $MSE_j(\hat{\Delta}_j)$?

$$b_\delta(\hat{\Delta}_1) = E[\bar{Y}-\bar{X}] - \delta = \overbrace{\mu_Y - \mu_X}^{\delta} - \delta = 0$$

• LINEARITY
• $\bar{X}, \bar{Y}$ UNBIASED

$$MSE_\delta(\hat{\Delta}_1) = \cancel{b_\delta^2(\hat{\Delta}_1)}^{0} + Var_\delta(\hat{\Delta}_1) = Var(\bar{Y}-\bar{X})$$

$$= Var(\bar{Y}) + Var(\bar{X}) = \frac{0.1}{5} + \frac{0.1}{3} = 0{,}0533$$

b) $b_\delta(\hat{\Delta}_2)$? $MSE_\delta(\hat{\Delta}_2)$?

$$b_\delta(\hat{\Delta}_2) = E[\hat{\Delta}_2] - \delta \underset{iid}{=} (3\mu_Y - 2\mu_X + \mu_X - 2\mu_X) - \delta$$

$$= \delta - \delta = 0 \Rightarrow \hat{\Delta}_2 \text{ unbiased}$$

Since unbiased:

$$MSE_\delta(\hat{\Delta}_2) = Var_\delta(\hat{\Delta}_2) \underset{\perp\!\!\!\perp}{=} 5\,Var(Y) + 3\,Var(X) = 5 \cdot 0{,}1 + 3 \cdot 0{,}1 = 0{,}8$$

c) In terms of mean square error, $\hat{\Delta}_1$ is better since $MSE_\delta(\hat{\Delta}_1) < MSE_\delta(\hat{\Delta}_2$

d) Point Estimator $\hat{\Delta}_1$ given $\{x_1, x_2, x_3\} = \{0{,}487, 0{,}565, 0{,}479\}$

$\{y_1, y_2, y_3, y_4, y_5\} = \{3.543, 3.284, 3.405, 3{,}551, 3.347\}$

$\hat{\Delta}_1 = \bar{Y} - \bar{X}$ Estimator for δ

$\hat{\delta}_1 = \bar{y} - \bar{x} = 2,9284$ Estimate for δ

Exercise 6,5

$X \rightarrow$ performance of our plant

$X_1, \dots, X_n \overset{iid}{\sim} N(\mu, 2)$

$\{x_1, \dots, x_5\} = \{91.60, 88.75, 90.80, 89.95, 91.30\}$

a) $CI_{95\%}(\mu)$? Width?

Confidence, in stead using one estimator we use two so we have a probability of θ being inside this interval is $1-\alpha$.

$X_1, \dots, X_n \overset{iid}{\sim} N$ and σ^2 known

$$CI_{1-\alpha}(\mu) = \left(\bar{x}_n - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x}_n + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

$\alpha = 0,05 \rightarrow z_{0,975} = 1,96$

$\bar{x}_5 = 90,48$

$\sigma = \sqrt{2}, n = 5$

$CI_{95\%}(\mu) = (89,24039, 91,71961)$

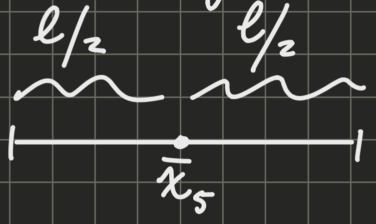
length of CI = 2,47922

$P\{\underbrace{89,\dots < \mu < 91,\dots}\} \neq 1-\alpha$

↳ we cannot say probability, since there is no random variable just numbers, so we have to say confidence not probability.

b)

↳ At 95% confidence error, what is the max. error made by estimating μ with sample mean.



For the sample mean, the max error possible is half the length if the real value is at the extremes

$\Rightarrow \frac{\text{length}}{2} \cong 1{,}24$

c) How many NEW measurements, $CI_{0,95}(\mu)$ with length at most 2,2?

$$\text{length} = 2 \cdot z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq 2.2 \Rightarrow \sqrt{n} > 2{,}519871$$

$$\Rightarrow n \geq 6{,}35 \Rightarrow n = 7$$

$\Rightarrow$ we need 2 new measurements.

Exercise 6,6

X : weight of flour bag (kg)

$X \sim N(\mu, 0{,}5^2)$

$\bar{x}_{100} = 15{,}3$

a) $CI_{0,90}(\mu)$? MAX ERROR by using sample mean?

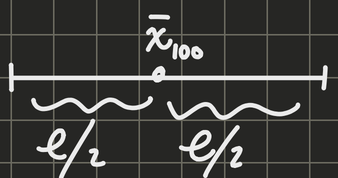
$X_1, \dots, X_{100} \overset{iid}{\sim} N(\mu, 0{,}5^2)$ with known variance

$$CI_{1-\alpha}(\mu) = \left(\bar{x}_n - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}},\ \bar{x}_n + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

$$\left.\begin{array}{l} \bar{x}_{100} = 15{,}3 \quad , \alpha = 100 \Rightarrow 1 - \frac{\alpha}{2} = 0{,}95 \\ z_{0{,}95} = 1{,}645 \quad \sigma = 0{,}5 \end{array}\right\}$$

$$\Rightarrow CI_{0{,}90}(\mu) = (15.21775,\ 15{,}38225)$$

At 90% level:



$$\Rightarrow \frac{15{,}38225 - 15{,}21775}{2} = 0{,}08225$$

b) n? : length $< 0{,}1$

$$\text{length} = 2 \cdot z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < 0{,}1$$

$$\Rightarrow n > 270.6 \Rightarrow n \geq 271$$

c) $X_1, \dots, X_n \overset{iid}{\sim} \Rightarrow$ CLT holds

$n = 100 \geq 50$

$$\bar{X}_n \overset{\text{approx}}{\sim} N\left(\mu, \frac{0{,}5^2}{100}\right)$$

$$\text{PIVOT}: \quad \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

No, we don't since CLT holds.

Exercise 6,7

X: # thumbtacks in a box

$X \sim f$, f unknown

$\bar{x}_{100} = 95.3$, $\sigma = 18,25$

μ = population mean

a) $CI_{0,95}(\mu)$?

No gaussianity assumption, so we need to check if CLT applies.

$n = 100 \geq 50$ $\quad X_1, \dots, X_n \overset{iid}{\sim} \Rightarrow \bar{X}_n \overset{approx}{\sim} N\left(\mu, \frac{(18,25)^2}{100}\right)$

$$CI_{1-\alpha} = \left(\bar{x}_{100} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} , \bar{x}_n + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

$$CI_{0,95}(\mu) = (91.723, 98.877)$$

b) n: length ≤ 4

$$2 \cdot z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq 4$$

$$2 \cdot 1,96 \cdot \frac{18,25}{\sqrt{n}} \leq 4$$

$n > 319,87 \Rightarrow n \geq 320$