13.01.2025

1. (**The 3σ Rule**). For any random variable X, most of the values of X lie within 3 standard deviations away from the mean.

X rand variable

$\mu = E(x)$

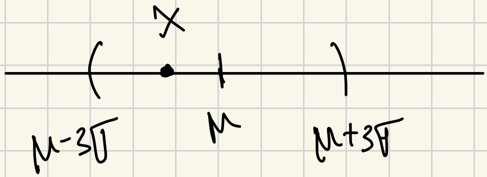
$\hookrightarrow$ expected value

$\sigma = \sqrt{V(x)}$

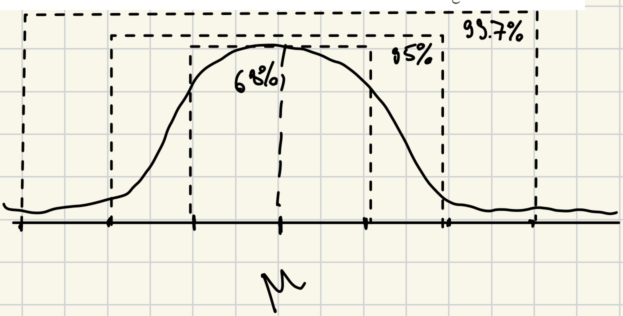
$V(x)$ $\hookrightarrow$ variance

σ $\hookrightarrow$ standard deviation

$$P(X \in (\mu - 3\sigma, \mu + 3\sigma)) \geq ?$$



Chebyshev's Inequality: $P(|X - E(X)| \geq \varepsilon) \leq \dfrac{V(X)}{\varepsilon^2}, \forall \varepsilon > 0.$



$X \sim N(\mu, \sigma)$

$$= P(-3\sigma < X - \mu < 3\sigma)$$

$$= P(|x - E(x)| < 3\sigma)$$

$$= 1 - P(|x - E(x)| \geq 3\sigma)$$

$$\Rightarrow \; P(|X - E(x)| \geq \underbrace{3\sigma}_{>0}) \geq 1 - \frac{V(x)}{(3\sigma)^2} = 1 - \frac{\cancel{V(x)}}{9\cancel{\sigma^2}} = 1 - \frac{1}{9} = \frac{8}{9} = 0.(8)$$

2. True or False: There is at least a 90% chance of the following happening: when flipping a coin 1000 times, the number of "heads" that appear is between 450 and 550.

X = no. of heads out of 1000 flips.

$P(450 < X < 550) \geq 90\%$ (True / False)

Chebyshev's Inequality: $P(|X - E(X)| \geq \varepsilon) \leq \dfrac{V(X)}{\varepsilon^2}, \forall \varepsilon > 0.$

$E(x) = 500 = n \cdot p$ also from the Binomial Distribution ↳ independent trails

ε

$-50 < X - 500 < 50$

$|X - 500| < 50$

$P(|X - 500| < 50) = 1 - P(|X - 500| \geq 50)$

$P(|X - 500| \geq 50) \geq 1 - \dfrac{V(X)}{50^2} = 1 - \dfrac{1}{10} = \dfrac{9}{10}$

$V(x) = n \cdot p \cdot q = 1000 \cdot (0.5) \cdot (0.5) = 250$

Binomial distribution formula; no. of trails; prob of success; prob of failure

Table solution

$P(450 < X < 500) \geq 80\%$ (?)

$X \in Bino(n=1000, p=\frac{1}{2})$

$E(X) = n \cdot p$, $V(X) = n \cdot p \cdot (1-p)$

Cheb. ineq. $\Rightarrow P(|X-500| \geq \varepsilon) \leq \frac{250}{\varepsilon^2} \Leftrightarrow$

$\Leftrightarrow P(|X-500| < \varepsilon) \geq 1 - \frac{250}{\varepsilon^2}$

we need to compute 'ε'

$P(-\varepsilon + 500 < X < \varepsilon + 500) \geq 1 - \frac{250}{\varepsilon^2}$

$\varepsilon = 50$

$P(450 < X < 550) \geq 1 - \frac{250}{2500} = \frac{9}{10} = 90\%$

Law of Large Numbers

$X_1, \dots, X_n$; rand X.

$E(X) = \mu$

$\frac{1}{n}(X_1 + \dots + X_n) \longrightarrow \mu, \; n \to \infty$

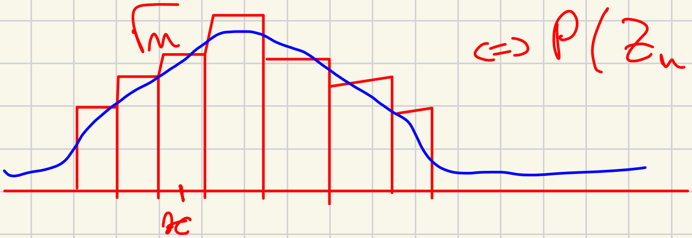
functions, not numbers

Central Limit Theorem

$S_n = X_1 + \dots + X_n$

$\sigma = \sqrt{V(X)}$

$$\frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} \longrightarrow Z \in N(0,1), \; n \to \infty$$



$\iff P(Z_n \leq x) \to P(Z \leq x),$
$n \to \infty, \forall x \in \mathbb{R}$

3. Installation of some software package requires downloading 82 files. On the average, it takes 15 sec to download a file, with a variance of 16 $\sec^2$. What is the probability that the software is installed in less than 20 minutes?

82 files

X = time to download a file

standard deviation

$E(X) = 15, \quad V(X) = 16 \Rightarrow \sigma = 4$

X_i = time for file 'i', $i = \overline{1, 82}$

$X_1, \dots, X_{82}$ independent + identically distributed

$P(S_n < 20 \text{ min}) \approx ?$

Central Limit Theorem(CLT) Let $X_1, \dots, X_n$ be independent random variables with the same expectation $\mu = E(X_i)$ and same standard deviation $\sigma = \sigma(X_i) = \mathrm{Std}(X_i)$ and let $S_n = \sum_{i=1}^{n} X_i$. Then, as $n \to \infty$,

$$Z_n = \frac{S_n - E(S_n)}{\sigma(S_n)} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \longrightarrow Z \in N(0,1), \text{ in distribution (in cdf), i.e. } F_{Z_n} \to F_Z = \Phi.$$

20 min = 20·60 sec = 1200 seconds.

$P(S_{82} < 1200) =$

$$= P\left(\underbrace{\frac{\frac{S_{82}}{82} - 15}{\frac{4}{\sqrt{82}}}}_{Z_{82}} < \frac{\frac{1200}{82} - 15}{\frac{4}{\sqrt{82}}} \right) =$$

$$\approx P\left(Z_{82} < -0.82\right) \overset{C.L.T.}{\approx} \text{normcdf}(-0.82) \approx$$

$$\approx 20\%$$

4. A sample of 3 observations, $X_1 = 0.4, X_2 = 0.7, X_3 = 0.9$, is collected from a continuous distribution with pdf

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1}, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases},$$

with $\theta > 0$, unknown. Estimate θ by the method of moments and by the method of maximum likelihood.

$x_1 = 0.4,\ x_2 = 0.7,\ x_3 = 0.9$ values of X with p.d.f.

$$f(x,\theta) = \begin{cases} \theta x^{\theta-1}, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

$\theta > 0$ unknown

Estimate θ with: • method of moments

• method of max. likelihood

Sol: 1) method of moments

$$E(X) = \bar{x} \Leftrightarrow \int_{-\infty}^{\infty} x \cdot f(x,\theta)\, dx = \frac{x_1 + x_2 + x_3}{3}$$

$$\int_0^1 x \cdot \theta \cdot x^{\theta-1}\, dx = \frac{2}{3}$$

$$\theta \cdot \frac{x^{\theta+1}}{\theta+1}\Bigg|_0^1 = \frac{2}{3} \Rightarrow \frac{\theta}{\theta+1} = \frac{2}{3} \Rightarrow \theta = 2$$

2) method of max. likelihood.

$$1)\ L(x_1, x_2, x_3, \theta) = \prod_{i=1}^{3} f(x_i;\theta) =$$

$$= \theta^3 \cdot (x_1 \cdot x_2 \cdot x_3)^{\theta - 1} \Leftrightarrow$$

$$\Leftrightarrow 2)\ \ln\left(L(x_1, x_2, x_3, \theta)\right) = 3\ln\theta + (\theta - 1)\cdot \ln(x_1 \cdot x_2 \cdot x_3)$$

$$\Leftrightarrow 3)\ \frac{\partial \ln L}{\partial \theta}(x_1, x_2, x_3, \theta) = \frac{3}{\theta} + \ln(x_1 \cdot x_2 \cdot x_3) = 0 \Leftrightarrow$$

$$\Leftrightarrow \theta = -\frac{3}{\ln(0.252)} \approx 2.17$$

5. A sample $X_1, \ldots, X_n$ is drawn from a distribution with pdf

$$f(x;\theta) = \frac{1}{2\theta} e^{-\frac{x}{2\theta}},\ x > 0$$

($\theta > 0$), which has mean $\mu = E(X) = 2\theta$ and variance $\sigma^2 = V(X) = 4\theta^2$. Find
a) the method of moments estimator, $\overline{\theta}$, for θ;
b) the efficiency of $\overline{\theta}$, $e(\overline{\theta})$;
c) an approximation for the standard error of the estimate in a), $\sigma_{\overline{\theta}}$, if the sum of 100 observations is 200.

$X_1, \ldots, X_n$ sample

$$f(x, \theta) = \begin{cases} \frac{1}{2\theta} \cdot e^{-\frac{x}{2\theta}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

$\mu = E(X) = 2\theta$

$\sigma^2 = V(X) = 4\theta^2$

a) $E(X) = \overline{X} = 2\theta \Rightarrow \overline{\theta} = \frac{\overline{X}}{2}$

↳ the estimator

b) efficiency (last page seminar)

$e(\overline{\theta}) = ?$; θ absolutely correct

$E(\overline{\theta}) = \theta$ and $V(\overline{\theta}) \to 0, n \to \infty$

$E(\overline{\theta}) = E\left(\frac{1}{2}\overline{X}\right) = \frac{1}{2} E\left(\frac{1}{n} \cdot (X_1 + \dots + X_n)\right) =$

$= \frac{1}{2n}\left(\underbrace{E(X_1)}_{2\theta} + \dots + \underbrace{E(X_n)}_{2\theta}\right) = \frac{1}{2} \cdot 2\theta = \theta$

$V(\theta) = V\left(\frac{1}{2n}(X_1 + \dots + X_n)\right) = \frac{1}{(2n)^2} \cdot \left(\underbrace{V(X_1)}_{4\theta^2} + \dots + \underbrace{V(X_n)}_{4\theta^2}\right)$

$= \frac{\theta^2}{n} \to 0, \ n \to \infty$

$$e(\bar{\theta}) = \frac{1}{J_n(\theta)\cdot V(\bar{\theta})} = \frac{1}{J_n(\theta)\cdot\frac{\theta^2}{n}}$$

$$J_n(\theta) = -E\left(\frac{\partial^2 \ln L}{\partial \theta^2}(X_1,\dots,X_n;\theta)\right) \Longrightarrow$$

$J_n(\theta) = n\cdot J_1(\theta)$ if the range of X does not depend on θ

$$\underbrace{P(X>0)}_{\text{range does not depend on } \theta.} = 100\%$$

$$\Longrightarrow J_n(\theta) = n\cdot\left(-E\left(\frac{\partial^2 \ln L}{\partial\theta^2}(X_1;\theta)\right)\right)$$

$$L(X_1;\theta) = f(X_1,\theta) = \frac{1}{2\theta}\cdot e^{-\frac{X_1}{2\theta}}$$

$$\ln L(X_1;\theta) = -\ln(2\theta) - \frac{X_1}{2\theta}$$

$$\frac{\partial \ln L}{\partial\theta}(X_i;\theta) = -\frac{1}{\cancel{2}\theta}\cdot\cancel{2} + \frac{X_1}{(2\theta)^2}\cdot 2\ ;$$

$$I_n(\theta) = n \cdot \left(-E\left(\frac{1}{\theta^2} - \frac{X_1}{\theta^3}\right)\right) = n\left(-\frac{1}{\theta^2} + \frac{1}{\theta^3}\underbrace{E(X_1)}_{2\theta}\right) =$$

$$= \frac{n}{\theta^2} \Rightarrow e(\bar{\theta}) = 1.$$