Seminar Nr. 7, Inequalities; Central Limit Theorem; Point Estimators

Theory Review

Markov's Inequality: $P(|X| \geq a) \leq \dfrac{1}{a} E(|X|), \forall a > 0.$

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

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 $S_n = \sum\limits_{i=1}^{n} X_i$. Then, as $n \to \infty$,

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

Point Estimators

- method of moments: solve the system $\nu_k = \overline{\nu}_k$, for as many parameters as needed ($k = 1, \dots,$ nr. of unknown parameters);

- method of maximum likelihood: solve $\dfrac{\partial \ln L(X_1, \dots, X_n; \theta)}{\partial \theta_j} = 0$, where $L(X_1, \dots, X_n; \theta) = \prod\limits_{i=1}^{n} f(X_i; \theta)$ is the likelihood function;

- **standard error** of an estimator $\overline{\theta}$: $\sigma_{\hat{\theta}} = \sigma(\overline{\theta}) = \sqrt{V(\overline{\theta})}$;

- **Fisher information** $I_n(\theta) = -E\left[\dfrac{\partial^2 \ln L(X_1, \dots, X_n; \theta)}{\partial \theta^2}\right]$; if the range of X does not depend on θ, then $I_n(\theta) = nI_1(\theta)$;

- **efficiency** of an absolutely correct estimator $\overline{\theta}$: $e(\overline{\theta}) = \dfrac{1}{I_n(\theta)V(\overline{\theta})}$.

- an estimator $\overline{\theta}$ for the target parameter θ is
  - **unbiased**, if $E(\overline{\theta}) = \theta$;
  - **absolutely correct**, if $E(\overline{\theta}) = \theta$ and $V(\overline{\theta}) \to 0$, as $n \to \infty$;
  - **MVUE** (minimum variance unbiased estimator), if $E(\overline{\theta}) = \theta$ and $V(\overline{\theta}) \leq V(\hat{\theta}),\ \forall \hat{\theta}$ unbiased estimator;
  - **efficient**, if $e(\overline{\theta}) = 1$.

- $\overline{\theta}$ efficient => $\overline{\theta}$ MVUE.

---

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

Solution:

Let X be a r.v. with mean $E(X) = \mu$ and standard deviation $\sigma(X) = \sqrt{V(X)} = \sigma$. In Chebyshev's inequality let $\varepsilon = k\sigma$, for $k = \overline{1,3}$. Then

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

For $k = 1$ that means

$$P(|X - \mu| < \sigma) \quad \geq \quad 0 \text{ (not much)}.$$

For $k = 2$, we have

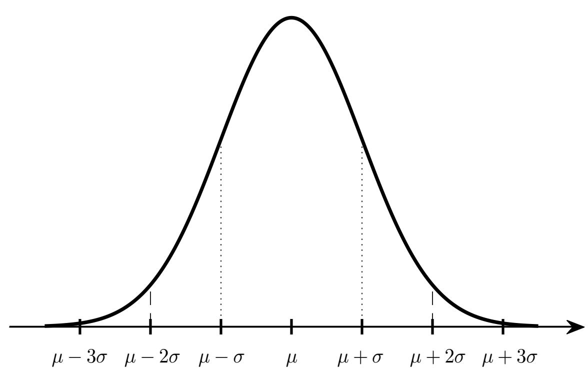
$$P(|X - \mu| < 2\sigma) \quad \geq \quad 1 - \frac{1}{4} \quad = \quad \frac{3}{4} \quad = \quad 0.75.$$

Finally, for $k = 3$, we get

$$P(|X - \mu| < 3\sigma) \;\geq\; 1 - \frac{1}{9} \;=\; \frac{8}{9} \;\approx\; 0.89.$$

The **3σ Rule** states that: most of the values (at least 89%) that a random variable takes, lie within 3 standard deviations (3σ) away from the mean.
Now, for *symmetric* distributions, these probabilities are *much* higher. For $X \in N(\mu, \sigma)$ (recall, for $X \in N(\mu, \sigma), E(X) = \mu, V(X) = \sigma^2$), we have:

$$\begin{aligned} P(|X - \mu| < \sigma) &\approx 0.68, \\ P(|X - \mu| < 2\sigma) &\approx 0.95, \\ P(|X - \mu| < 3\sigma) &\approx 0.99. \end{aligned}$$



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.

<u>Solution</u>:
Let X denote the number of "heads" that appear when tossing a coin 1000 times. Then $X \in Bino\left(1000, \frac{1}{2}\right)$, so

$$\begin{aligned} E(X) &= np = 500, \\ V(X) &= npq = 250. \end{aligned}$$

Now, the problem asks about the "chance" (i.e. probability) that this number is between 450 and 550, so about

$$P(450 < X < 550).$$

Chebyshev's inequality gives information about

$$P(|X - 500| < \varepsilon),$$

which we rewrite as

$$\begin{aligned} P(|X - 500| < \varepsilon) &= P(-\varepsilon < X - 500 < \varepsilon) \\ &= P(500 - \varepsilon < X < 500 + \varepsilon). \end{aligned}$$

So, we take $\varepsilon = 50$ in Chebyshev's inequality. Then we get

$$\begin{aligned} P(450 < X < 550) &= P(500 - 50 < X < 500 + 50) \\ &= P(-50 < X - 500 < 50) \\ &= P(|X - 500| < 50) \\ &\geq 1 - \frac{250}{\varepsilon^2} \\ &= 1 - \frac{250}{2500} \;=\; 0.9, \end{aligned}$$

so the statement is true.

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?

<u>Solution</u>:
Let X_i denote the time it takes to download file i. Then, for every $i = 1, \ldots, 82$,

$$\begin{aligned} \mu &= E(X_i) = 15 \text{ sec}, \\ \sigma &= \sqrt{V(X_i)} = \sqrt{16} = 4 \text{ sec}. \end{aligned}$$

The *entire* software is installed in

$$S_{82} = X_1 + X_2 + \ldots X_{82} \text{ sec}.$$

We have a sample of size $n = 82$, $X_1, X_2, \ldots, X_{82}$. Convert the time into seconds, $20 \text{ min} = 1200 \text{ sec}$. So, we want to compute

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

By the CLT, we have

$$\begin{aligned} P(S_{82} < 1200) &= P\left(\frac{S_{82} - n\mu}{\sigma\sqrt{n}} < \frac{1200 - n\mu}{\sigma\sqrt{n}}\right) \\ &= P\left(Z_n < \frac{1200 - 82 \cdot 15}{4 \cdot \sqrt{82}}\right) \\ &= P(Z_n < -0.8282) \overset{\text{CLT}}{\approx} P(Z < -0.8282) \\ &= F_Z(-0.8282) = \mathit{normcdf}(-0.8282) = 0.2038. \end{aligned}$$

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.

<u>Solution</u>:
<u>Method of moments</u>:
There is only one unknown, θ, so we solve the system

$$\nu_1 = \overline{\nu}_1,$$

where

$$\begin{aligned} \nu_1 &= E(X) = \int_{\mathbb{R}} x f(x)\, dx \\ &= \theta \int_0^1 x^\theta\, dx \\ &= \frac{\theta}{\theta+1} x^{\theta+1}\Big|_{x=0}^{x=1} = \frac{\theta}{\theta+1} \end{aligned}$$

and

$$\begin{aligned} \overline{\nu}_1 &= \overline{x} = \frac{x_1 + x_2 + x_3}{3} \\ &= \frac{0.4 + 0.7 + 0.9}{3} = \frac{2}{3}. \end{aligned}$$

Solve for θ

$$\begin{aligned} \frac{\theta}{\theta+1} &= \overline{x}, \\ \theta &= \overline{x}(\theta+1), \\ \theta(1-\overline{x}) &= \overline{x}, \end{aligned}$$

to get

$$\overline{\theta} \;=\; \frac{\overline{x}}{1-\overline{x}} \;=\; 2.$$

<u>Method of maximum likelihood</u>:
Again, having only one unknown, θ, we have only one equation

$$\frac{\partial \ln L(x_1,\ldots,x_n;\theta)}{\partial\theta} = 0.$$

The likelihood function is the joint density of the vector (X_1, X_2, X_3) :

$$\begin{aligned} L(x_1,x_2,x_3;\theta) &= \prod_{i=1}^{3} f(x_i;\theta) \\ &= \prod_{i=1}^{3}\left(\theta x_i^{\theta-1}\right) \;=\; \theta^3\left(\prod_{i=1}^{3} x_i\right)^{\theta-1} \\ \ln L &= 3\ln\theta + (\theta-1)\sum_{i=1}^{3}\ln x_i \\ \frac{\partial \ln L}{\partial\theta} &= \frac{3}{\theta} + \sum_{i=1}^{3}\ln x_i. \end{aligned}$$

Solve $\dfrac{\partial \ln L}{\partial\theta} = 0$ for θ, to find

$$\hat{\theta} \;=\; -\frac{3}{\displaystyle\sum_{i=1}^{3}\ln x_i} \;=\; 2.1766.$$

<u>Note</u>: In the case where the two estimators *do not* coincide, the MLE is more trustworthy.

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

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

($\theta>0$), 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.

<u>Solution:</u>
a) Again, there is only one unknown, θ, so we solve the system $\nu_1 = \overline{\nu}_1$, where

$$\nu_1 \;=\; E(X) \;=\; 2\theta$$

and

$$\overline{\nu}_1 \;=\; \overline{X} \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Solve for θ

$$2\theta = \overline{X},$$

to get

$$\overline{\theta} = \frac{1}{2}\overline{X}.$$

Note In this case, this is also the MLE.
Now, *efficiency* is only computed for *absolutely correct* estimators. Let us check the absolute correctness. We have

$$\begin{aligned} E(\overline{\theta}) &= E\left(\frac{1}{2}\overline{X}\right) &= \frac{1}{2}E(\overline{X}) \\ &= \frac{1}{2}\cdot\mu \;=\; \frac{1}{2}\cdot 2\theta &= \theta \end{aligned}$$

and

$$\begin{aligned} V(\overline{\theta}) &= V\left(\frac{1}{2}\overline{X}\right) = \frac{1}{4}V(\overline{X}) \\ &= \frac{1}{4}\cdot\frac{\sigma^2}{n} = \frac{1}{4n}4\theta^2 = \frac{\theta^2}{n} \overset{n\to\infty}{\longrightarrow} 0, \end{aligned}$$

so $\overline{\theta}$ is an absolutely correct estimator for θ.

b) The efficiency is given by

$$e(\overline{\theta}) = \frac{1}{I_n(\theta)V(\overline{\theta})},$$

where

$$I_n(\theta) = -E\left[\frac{\partial^2 \ln L(X_1,\ldots,X_n;\theta)}{\partial\theta^2}\right]$$

is the Fisher information.
Since the range of X *does not* depend on θ, we have

$$I_n(\theta) = nI_1(\theta),$$

where

$$I_1(\theta) = -E\left[\frac{\partial^2 \ln L(X_1;\theta)}{\partial\theta^2}\right].$$

We proceed with the computations:

$$\begin{aligned} L(X_1;\theta) &= \frac{1}{2\theta}e^{-\frac{1}{2\theta}X_1}, \\ \ln L(X_1;\theta) &= -\ln(2\theta) - \frac{1}{2\theta}X_1 = -\ln 2 - \ln\theta - \frac{1}{2\theta}X_1, \\ \frac{\partial \ln L}{\partial\theta} &= -\frac{1}{\theta} + \frac{1}{2}\cdot\frac{1}{\theta^2}X_1, \\ \frac{\partial^2 \ln L}{\partial\theta^2} &= \frac{1}{\theta^2} - \frac{1}{\theta^3}X_1, \\ -E\left[\frac{\partial^2 \ln L}{\partial\theta^2}\right] &= -\frac{1}{\theta^2} + \frac{1}{\theta^3}E(X_1) = -\frac{1}{\theta^2} + \frac{2}{\theta^2} = \frac{1}{\theta^2}. \end{aligned}$$

So

$$I_n(\theta) = \frac{n}{\theta^2} \text{ and } e(\overline{\theta}) = 1,$$

which means the estimator is *efficient* and, thus, also a MVUE.

c) The standard error is

$$\begin{aligned}\sigma_{\overline{\theta}} &= \sigma(\overline{\theta}) = \sqrt{V(\overline{\theta})} \\ &= \frac{\theta}{\sqrt{n}} \approx \frac{\overline{\theta}}{\sqrt{n}}.\end{aligned}$$

If the sum of 100 observations is 200, then $n = 100$ and $\overline{X} = \dfrac{200}{100} = 2$.
The estimator is

$$\overline{\theta} = 1$$

and the standard error is

$$\sigma_{\overline{\theta}} \approx \frac{1}{10} = 0.1.$$