

# Normal model: introduction

Let us consider the following Normal model, for which  $\mu$  and  $\sigma^2$  are parameters.

$$\left\{ N(\mu, \sigma^2); \mu \in \mathbb{R}, \sigma^2 > 0 \right\}$$

The empirical mean  $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$  is a natural estimator for  $\mu$ . As we have seen, it is the Maximum Likelihood Estimator (MLE) and the Method of Moments (MoM) estimator for  $\mu$ .

# Consistency of estimators (1/4)

**Consistency:** An estimator  $\hat{\theta}_n$  is said to be consistent for  $\theta$  if it converges in probability to the true value of the parameter. We have that  $\hat{\theta}_n \xrightarrow{P} \theta$  as  $n$  goes to  $\infty$ .

By the Law of Large Numbers (LLN), we have that

$$\bar{x}_n \xrightarrow{P} E[X] \quad \text{as } n \rightarrow \infty. \text{ Here, } E[X] = \mu$$

So  $\bar{x}_n$  is a **consistent** estimator for  $\mu$ .

**Continuous Mapping Theorem:** If we have  $\hat{\theta}_n \xrightarrow{P} \theta$  as  $n$  goes to  $\infty$ , then for a continuous function  $g(\hat{\theta}_n)$ , the convergence in probability is preserved. We have that  $g(\hat{\theta}_n) \xrightarrow{P} g(\theta)$  as  $n$  goes to  $\infty$ .

## Consistency of estimators (2/4)

So by the Continuous Mapping Theorem (CMT), here for  $g(x) = x^2$ , we have that

$$(\bar{x}_n)^2 \xrightarrow{\mathcal{P}} (E[X])^2 \quad \text{as } n \rightarrow \infty. \text{ Here, } (E[X])^2 = \mu^2$$

**Slutsky Theorem:** For two random variables  $X_n$  and  $Y_n$ , if  $X_n \xrightarrow{P} X$  as  $n$  goes to  $\infty$  and  $Y_n \rightarrow c$ , where  $c$  is a constant, then we have that  $X_n + Y_n \xrightarrow{P} X + c$  as  $n$  goes to  $\infty$ .

## Consistency of estimators (3/4)

We recall that the empirical variance is given by

$$\begin{aligned} S_n^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \\ &= \frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x}_n)^2 \end{aligned}$$

And since, by the LLN, we have that

$$\frac{1}{n} \sum_{i=1}^n x_i^2 \quad \xrightarrow{\mathcal{P}} \quad E[X^2] \quad \text{as } n \rightarrow \infty$$

## Consistency of estimators (4/4)

Therefore, by the Slutsky theorem, we have that

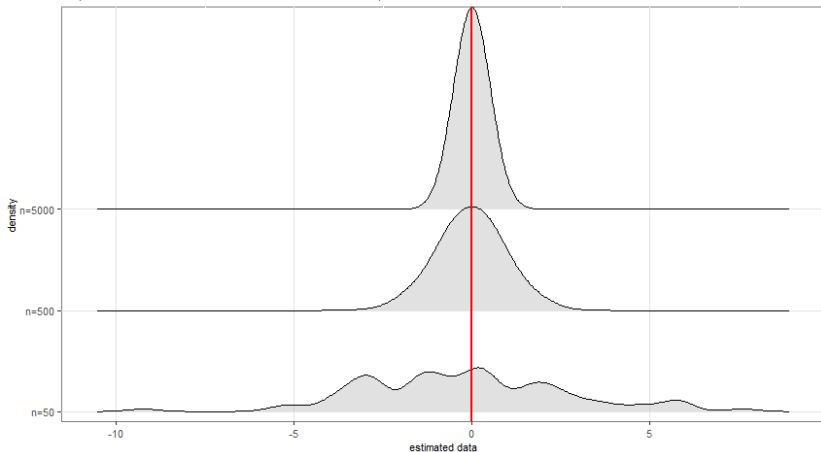
$$\frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x}_n)^2 \xrightarrow{\mathcal{P}} E[X^2] - (E[X])^2$$

So we conclude that  $S_n^2$  is a **consistent** estimator for  $\text{var}_\theta(X) = \sigma_\theta^2$ .

# Visualization of consistency

## Consistency of estimators

*Empirical mean - the standard deviation shrinks as the sample size increases*



# Bias of estimators (1/3)

We consider the empirical variance  $S_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$ . Is it an unbiased estimator for  $\sigma_\theta^2$  (denoted also equivalently  $\text{var}_\theta(X)$ ). Let us first rearrange its expression, and then compute its expectation.

$$\begin{aligned} S_n^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = \frac{1}{n} \sum_{i=1}^n (x_i - E_\theta[X] + E_\theta[X] - \bar{x}_n)^2 \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - E_\theta[X])^2 + \frac{1}{n} \sum_{i=1}^n (E_\theta[X] - \bar{x}_n)^2 + 2 \frac{1}{n} \sum_{i=1}^n (x_i - E_\theta[X]) (E_\theta[X] - \bar{x}_n) \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - E_\theta[X])^2 + (E_\theta[X] - \bar{x}_n)^2 + 2 (E_\theta[X] - \bar{x}_n) \underbrace{\frac{1}{n} \sum_{i=1}^n (x_i - E_\theta[X])}_{(\bar{x}_n - E_\theta[X])} \\ &\quad \underbrace{\phantom{\frac{1}{n} \sum_{i=1}^n (x_i - E_\theta[X])^2} + (E_\theta[X] - \bar{x}_n)^2 + 2 (E_\theta[X] - \bar{x}_n) (\bar{x}_n - E_\theta[X])}_{(E_\theta[X] - \bar{x}_n)^2 - 2 (E_\theta[X] - \bar{x}_n)^2} \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - E_\theta[X])^2 - (E_\theta[X] - \bar{x}_n)^2 \end{aligned}$$

## Bias of estimators (2/3)

$$\begin{aligned}E_{\theta}[S_n^2] &= E_{\theta}\left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2\right] \\&= E_{\theta}\left[\frac{1}{n} \sum_{i=1}^n (x_i - E_{\theta}[X])^2 - (E_{\theta}[X] - \bar{x}_n)^2\right] \\&= E_{\theta}\left[\frac{1}{n} \sum_{i=1}^n (x_i - E_{\theta}[X])^2\right] - E_{\theta}\left[(E_{\theta}[X] - \bar{x}_n)^2\right] \\&= \frac{1}{n} \sum_{i=1}^n \underbrace{E_{\theta}\left[(x_i - E_{\theta}[X])^2\right]}_{\text{var}_{\theta}(X)} - \underbrace{E_{\theta}\left[(E_{\theta}[X] - \bar{x}_n)^2\right]}_{\text{var}_{\theta}(\bar{x}_n)} \\&= \sigma_{\theta}^2 - \text{var}_{\theta}(\bar{x}_n) \\&= \sigma_{\theta}^2 - \frac{\sigma_{\theta}^2}{n} \\&= \left(\frac{n-1}{n}\right) \sigma_{\theta}^2 \quad \left( \neq \sigma_{\theta}^2 \text{ for all } n \right).\end{aligned}$$



## Bias of estimators (3/3)

Therefore we conclude that  $S_n^2$  is a **biased estimator** for  $\text{var}_\theta(X) = \sigma_\theta^2$ . But since  $E_\theta[S_n^2] = \left(\frac{n-1}{n}\right) \sigma_\theta^2 \xrightarrow{n \rightarrow \infty} \sigma_\theta^2$

$S_n^2$  is an **asymptotically unbiased** estimator for  $\text{var}_\theta(X) = \sigma_\theta^2$ . Now we consider  $S_n^{2*} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2$ . We have

$$\begin{aligned} E_\theta[S_n^{2*}] &= E_\theta \left[ \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \right] = \frac{n}{n-1} E_\theta \left[ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \right] \\ &= \frac{n}{n-1} E_\theta \left[ \underbrace{\sum_{i=1}^n (x_i - \bar{x}_n)^2}_{\text{var}_\theta(X)} \right] = \frac{n}{n-1} E_\theta[S_n^2] = \frac{n}{n-1} \left( \frac{n-1}{n} \right) \sigma_\theta^2 = \sigma_\theta^2 \end{aligned}$$

So  $S_n^{2*}$  is an **unbiased** estimator for  $\text{var}_\theta(X) = \sigma_\theta^2$ .

# Efficiency of estimators (1/)

The **mean square error** (MSE), is defined as

$$R(T(\mathbf{x}), \theta) = E_{\theta} \left[ \left( T(\mathbf{x}) - g(\theta) \right)^2 \right] = \text{var}_{\theta}(T(\mathbf{x})) + b_T^2(\theta)$$

$$\begin{aligned} \text{MSE} = R(T(\mathbf{x}), \theta) &= E_{\theta} \left[ \left( T(\mathbf{x}) - g(\theta) \right)^2 \right] \\ &= E_{\theta} \left[ \left( T(\mathbf{x}) - E_{\theta}[T(\mathbf{x})] + E_{\theta}[T(\mathbf{x})] - g(\theta) \right)^2 \right] \\ &= E_{\theta} \left[ \underbrace{\left( T(\mathbf{x}) - E_{\theta}[T(\mathbf{x})] \right)^2}_{\text{var}_{\theta}(T(\mathbf{x}))} \right] + E_{\theta} \left[ \underbrace{\left( E_{\theta}[T(\mathbf{x})] - g(\theta) \right)^2}_{\text{constant}} \right] \\ &\quad + 2 E_{\theta} \left[ \underbrace{\left( T(\mathbf{x}) - E_{\theta}[T(\mathbf{x})] \right)}_{=0} \right] + E_{\theta} \left[ \underbrace{\left( E_{\theta}[T(\mathbf{x})] - g(\theta) \right)}_{\text{constant}} \right] \\ &= \text{var}_{\theta}(T(\mathbf{x})) + \left( E_{\theta} \left[ T(\mathbf{x}) - g(\theta) \right] \right)^2 + 0 \\ &= \text{var}_{\theta}(T(\mathbf{x})) + b_T^2(\theta) \end{aligned}$$

## Efficiency of estimators (2/5)

The MSE for the empirical mean is therefore

$$\begin{aligned}MSE &= R(\bar{x}, \theta) = \text{var}_{\theta}(\bar{x}) + b_{\bar{x}}^2(\theta) \\&= \frac{\text{var}_{\theta}(\bar{x})}{n} + 0 \\&= \frac{\sigma_{\theta}^2}{n}\end{aligned}$$

We notice that it decreases to 0 as  $n$  gets large (a MSE should always decrease to 0).

$$\frac{\sigma_{\theta}^2}{n} \xrightarrow{n \rightarrow \infty} 0$$

## Efficiency of estimators (3/5)

The **Fisher information** for one observation is denoted  $I(\theta)$ . The Fisher information for the sample  $x_1, x_2, \dots, x_n$  is denoted  $I_n(\theta)$ . It is defined as follow

**definition I :**  $var_{\theta}(S(x, \theta)) = E[S^2(x, \theta)]$

**definition II:**  $-E\left[\frac{\partial^2 \ln \mathcal{L}(\theta | \mathbf{x})}{\partial \theta^2}\right] = -E[l'']$

For the model (i), we first recall the expression for the second order derivative

$$\frac{\partial^2 \ln \mathcal{L}(\mu | \mathbf{x})}{\partial \mu^2} = \frac{1}{\sigma^2} (-n) = -\frac{n}{\sigma^2}$$

## Efficiency of estimators (4/5)

As a consequence, we have that (using the first definition I)

$$I(\mu) = \text{var}_{\mu}(S(x, \mu)) = E[S^2(x, \mu)] = E\left[\frac{(x, \mu)^2}{(\sigma^2)^2}\right] = \sigma^4 \underbrace{E[(x, \mu)^2]}_{\text{var}(x) = \sigma^2} = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}$$

Finally,  $I_n(\mu) = n I(\mu) = n/\sigma^2$ .

Or equivalently, using the definition II

$$I_n(\mu) = \text{var}_{\mu}(S(x, \mu)) = - \underbrace{E\left[\frac{\partial^2 \ln \mathcal{L}(\mu \mid \mathbf{x})}{\partial \mu^2}\right]}_{=n \frac{1}{\sigma^2}} = - \left( - \frac{n}{\sigma^2} \right) = \frac{n}{\sigma^2}$$

## Efficiency of estimators (5/5)

An estimator is said to be **efficient** if it attains the **Cramér-Rao bound** (CRB), defined as follow

$$CRB = \frac{(g'(\theta))^2}{I_n(\theta)}$$

where  $I_n(\theta)$  is the **Fisher information** of the sample. Moreover, since we are dealing here with the identity link function for the parameter, i.e.  $g(\mu) = \mu$ , we have  $g'(\mu) = (\mu)' = 1$ .

We obtain the following expression for the Cramér-Rao bound

$$CRB = \frac{1}{\frac{n}{\sigma^2}} = \frac{g'(\mu)}{I_n(\mu)} = \frac{\sigma^2}{n} \quad \text{or, for 1 observation } CRB = \sigma^2.$$

We conclude that our estimator  $\bar{x}$  is an efficient estimator for  $\mu$ . 14 / 15

# References

Course notes

The R Project for Statistical Computing:  
<https://www.r-project.org/>