How good are estimators and how to find them?

Given two estimators $\hat{\theta}_1$, $\hat{\theta}_2$,

We can measure "goodness" by.

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

<u>Relative efficiency.</u>

How do two estimators compare?

<u>Def</u> relative efficiency of $\hat{\theta}_1$, $\hat{\theta}_2$

$$\mathrm{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{MSE(\hat{\theta}_2)}{MSE(\hat{\theta}_1)}.$$

notation

<u>Note:</u> for $\hat{\theta}_1$, $\hat{\theta}_2$ are unbiased, then

$$\mathrm{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)}.$$

$\hat{\theta}_2$ is better

If

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) < 1 \iff MSE(\hat{\theta}_2) < MSE(\hat{\theta}_1)$$

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) > 1 \iff MSE(\hat{\theta}_2) > MSE(\hat{\theta}_1)$$

$\hat{\theta}_1$ is better

Example (HW7).

$X_1, X_2, \ldots, X_n$ sampled from $N(\theta, \sigma^2)$

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}, \quad \hat{\theta}_2 = \frac{X_1 + \bar{X}}{2}$$

Then

$$MSE(\hat{\theta}_1) = \frac{\sigma^2}{n}, \quad MSE(\hat{\theta}_2) = \frac{\sigma^2(n+3)}{4n}$$

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\frac{\cancel{\sigma^2}(n+3)}{4\cancel{n}}}{\cancel{\sigma^2}/\cancel{n}} = \frac{n+3}{4} > 1$$

$= 1$ if $n=1$.

> 1 if $n \geq 2$.

therefore $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$.

Consistency

Want to understand how an estimator gets better as you take larger samples.

Def A sequence of estimators $\hat{\theta}_n$ is consistent for θ as $n \to \infty$ if.

$$P(|\hat{\theta}_n - \theta| > \varepsilon) \longrightarrow 0 \quad \text{as} \quad n \to \infty.$$

for any $\varepsilon > 0$.

known as convergence in probability.

Recall $X_n \to X$ in probability as $n \to \infty$

if $P(|X_n - X| > \varepsilon) \longrightarrow 0$ as $n \to \infty$.

for any $\varepsilon > 0$

$\hat{\theta}_n$ is consistent with θ if $\hat{\theta}_n \to \theta$ in probability as $n \to \infty$.

How to show consistency?

Theorem Let $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n$ be a sequence of estimators for θ. Then

$$P(|\hat{\theta}_n - \theta| > \varepsilon) \leq \frac{MSE(\hat{\theta}_n)}{\varepsilon^2}$$

Chebyshev.

Further if $MSE(\hat{\theta}_n) \to 0$ as $n \to \infty$

then $\hat{\theta}_n$ is consistent for θ.

One way to check for consistency is to check that $MSE(\hat{\theta}_n) \to 0$.

Example

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \quad , \quad \hat{\theta}_2 = \frac{X_1 + \bar{X}}{2}.$$

$N(\theta, \sigma^2)$

$MSE(\hat{\theta}_1) = \frac{\sigma^2}{n}$, $MSE(\hat{\theta}_2) = \frac{\sigma^2(n+3)}{4n}$

$\to 0$ $n \to \infty$. $\frac{\sigma^2(n+3)}{4n} \to \frac{\sigma^2}{4} \neq 0$

$\Rightarrow$ $\hat{\theta}_1$ is consistent.

↑ not proof that $\hat{\theta}_2$ is not consistent.

How to show not consistent?

Properties of convergence in probability.

$\hat{\theta}_n \to \theta$, $\hat{\theta}_n' \to \theta'$ in probability.

① $\hat{\theta}_n + \hat{\theta}_n' \to \theta + \theta'$ in probability.

② $\hat{\theta}_n \cdot \hat{\theta}_n' \to \theta \cdot \theta'$ in probability.

③ $\hat{\theta}_n / \hat{\theta}_n' \to \theta / \theta'_{\neq 0}$ in probability if $\theta' \neq 0$.

④ $g(\hat{\theta}_n) \to g(\theta)$ in probability

if g is continuous at θ.

Ex

$$\hat{\Theta}_n = \frac{X_1 + \bar{X}_n}{2}$$

sample size

by weak LLN.

$\bar{X}_n \to \Theta$ in probability.

$$\hat{\Theta}_n \to \frac{X_1 + \Theta}{2} \quad \text{in probability.}$$

not consistent.

Example $X_1, X_2, \ldots, X_n$ are a random sample from the distribution with pdf

$$f(x) = \begin{cases} \Theta e^{-\Theta x}, & x \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

exponential(Θ) $\quad \Theta > 0$.

What is a good estimator for Θ?

$$\mathbb{E} X_i = 1/\Theta \Rightarrow \Theta = 1/\text{mean}.$$

Good idea is to take

$$\hat{\theta}_n = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} X_i}$$

Not easy to calculate the Bias or MSE of $\hat{\theta}_n$.

How to show consistency of $\hat{\theta}_n$?

by the weak LLN.

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \frac{1}{\theta} \text{ in probability.}$$

since

$$g(x) = \frac{1}{x} \text{ is continuous at } \frac{1}{\theta} \neq 0.$$

then

$$\hat{\theta}_n = g\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \to g\left(\frac{1}{\theta}\right) = \frac{1}{1/\theta} = \theta.$$

in probability.

Therefore it is consistent.

Sufficient statistics

As we saw in the last example

$$\bar{X} = 1/n \sum_{i=1}^{n} X_i \quad \text{was enough}$$

information to estimate $\theta_0 = 1/\beta_0$.

↳ $\bar{X} = 1/n \sum_{i=1}^{n} X_i$ is sufficient.

Def

$U = u(X_1, X_2, \dots, X_n)$ is a sufficient statistic for θ_0 if the joint distribution of $(X_1, \dots, X_n) = (x_1, x_2, \dots, x_n)$ given $U = u$ doesn't depend on θ.

↑ knowing this is as much as you need to know to estimate θ.

Example

pdf $\quad f(x) = \begin{cases} \theta e^{-\theta x} & x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$

$X_1, X_2, \ldots, X_n$ a random sample

the joint pdf is.

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i) = f(x_1) \ldots f(x_n)$$

$$= \left[\theta e^{-\theta x_1}\right]\left[\theta e^{-\theta x_2}\right] \ldots \left[\theta e^{-\theta x_n}\right]$$

$$= \theta^n e^{-\theta(x_1 + x_2 + \ldots + x_n)}.$$

$$= \theta^n e^{-\theta n \bar{x}}, \quad \bar{x} = 1/n \sum_{i=1}^{n} x_i$$

sample mean

this implies that $\bar{x}$ is sufficient to estimate θ.

Likelihood function

Change notation.

Denote distribution

NOT! Conditional prob.

Continuous case $f(x \mid \theta)$ - pdf.

Discrete case $p(x \mid \theta)$

way of writing dependence on parameter

Def

Given $X_1, X_2, \ldots X_n$ a random sample from $f(x \mid \theta)$ or $p(x \mid \theta)$.

numbers.

The likelihood of $(x_1, x_2, \ldots x_n)$ is

$$L(x_1, x_2, \ldots, x_n \mid \theta) = \begin{cases} f(x_1, x_2, \ldots x_n \mid \theta) & \text{cont} \\ \text{or} & \\ p(x_1, x_2, \ldots, x_n \mid \theta) & \text{discrete} \end{cases}$$

Factorization criterion

Theorem Let $X_1, X_2, \ldots, X_n$ be a random sample from $f(x|\theta)$ or $p(x|\theta)$, then

$U = U(X_1, X_2, \ldots X_n)$ is a sufficient statistic for θ if.

$$L(x_1, x_2, \ldots, x_n | \theta) = g(u, \theta)\, h(x_1, x_2, \ldots, x_n)$$

where $u = U(x_1, x_2, \ldots, x_n)$

and $h(x_1, x_2, \ldots, x_n)$ doesn't depend on θ.

Example

$$f(x|\theta) = \begin{cases} \theta e^{-\theta x} & x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$L(x_1, x_2, \ldots, x_n | \theta) = \prod_{i=1}^{n} f(x_i | \theta).$$

$$= \theta^n e^{-\theta n \bar{x}}$$

$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ is a sufficient statistic for θ.

$$L(x_1, x_2, \ldots, x_n | \theta) = \underbrace{g(\bar{x}, \theta)}_{\theta^n e^{-\theta n \bar{x}}} \underbrace{h(x_1, \ldots, x_n)}_{1}.$$