Estimation: introduction

Let us consider that we observe some data $x_1, ..., x_n$ and we assume the following parametric Normal model, for which μ and σ^2 are parameters that we try to estimate.

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

The empirical mean $\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ is a natural estimator for μ. It is the Maximum Likelihood Estimator (MLE) and the Method of Moments (MoM) estimator for μ. The **parameters** of a model are estimated by **estimators**, which might have different properties or asymptotic properties. Examples of estimators: a mean, a variance, a median, the OLS estimator in linear regression, a Maximum likelihood estimator, an histogram (nonparametric), a best linear predictor (for example in time series), the Hovitz-Thompson estimator (sampling theory)...

Consistency of estimators (1/4)

Consistency: An estimator $\hat{\theta}_n$ is said to be consistent for θ 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 ∞.
By the Law of Large Numbers (LLN), we have that

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

So $\overline{x}_n$ is a **consistent** estimator for μ.

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

Consistency of estimators (2/4)

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

$$(\overline{x}_n)^2 \quad \xrightarrow{\mathcal{P}} \quad \left(E[X]\right)^2 \qquad \text{as } n \to \infty. \text{ Here, } \left(E[X]\right)^2 = \mu^2$$

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

Consistency of estimators (3/4)

We recall that the empirical variance is given by

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

And since, by the LLN, we have that

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

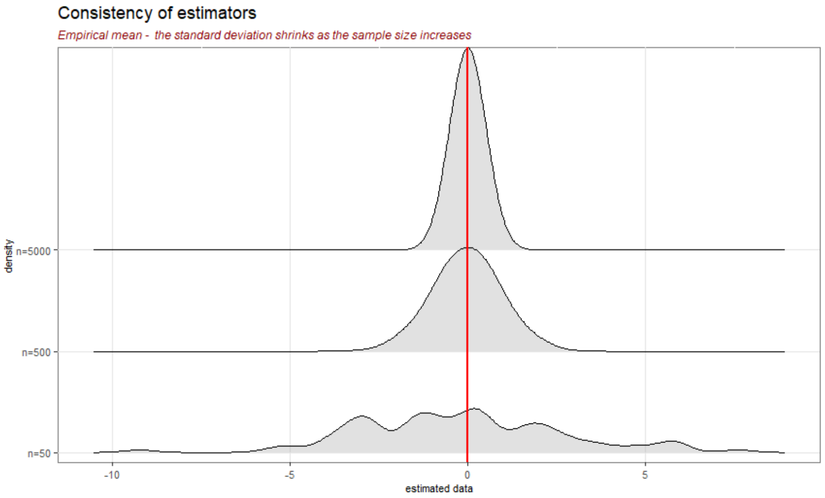
Consistency of estimators (4/4)

Therefore, by the Slutsky theorem, we have that

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

So we conclude that S_n^2 is a **consistent** estimator for $var_\theta(X) = \sigma_\theta^2$.

Visualization of consistency



Bias of estimators (1/3)

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

$$
\begin{aligned}
S_n^2 &= \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x}_n)^2 = \frac{1}{n}\sum_{i=1}^n \Big(x_i - E_\theta[X] + E_\theta[X] - \overline{x}_n\Big)^2 \\
&= \frac{1}{n}\sum_{i=1}^n \Big(x_i - E_\theta[X]\Big)^2 + \frac{1}{n}\sum_{i=1}^n \Big(E_\theta[X] - \overline{x}_n\Big)^2 + 2\frac{1}{n}\sum_{i=1}^n \Big(x_i - E_\theta[X]\Big)\Big(E_\theta[X] - \overline{x}_n\Big) \\
&= \frac{1}{n}\sum_{i=1}^n \Big(x_i - E_\theta[X]\Big)^2 + \underbrace{\Big(E_\theta[X] - \overline{x}_n\Big)^2 + 2\Big(E_\theta[X] - \overline{x}_n\Big)\underbrace{\frac{1}{n}\sum_{i=1}^n \Big(x_i - E_\theta[X]\Big)}_{\left(\overline{x}_n - E_\theta[X]\right)}}_{\left(E_\theta[X] - \overline{x}_n\right)^2 - 2\left(E_\theta[X] - \overline{x}_n\right)^2} \\
&= \frac{1}{n}\sum \Big(x_i - E_\theta[X]\Big)^2 - \Big(E_\theta[X] - \overline{x}_n\Big)^2
\end{aligned}
$$

Bias of estimators (2/3)

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

Bias of estimators (3/3)

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

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

$$E_\theta[S_n^{2*}] = E_\theta\left[\frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x}_n)^2\right] = \frac{n}{n-1}\ E_\theta\left[\frac{1}{n}\sum_{i=1}^n (x_i - \overline{x}_n)^2\right]$$

$$= \frac{n}{n-1}\ \underbrace{\underbrace{E_\theta\left[\sum_{i=1}^n (x_i - \overline{x}_n)^2\right]}_{var_\theta(X)}}_{S_n^2} = \frac{n}{n-1}\ E_\theta[S_n^2] = \frac{n}{n-1}\left(\frac{n-1}{n}\right)\sigma_\theta^2 \ = \ \sigma_\theta^2$$

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

Efficiency of estimators (1/5)

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

$$R(T(\boldsymbol{x}),\theta) = E_\theta\left[\Big(T(\boldsymbol{x}) - g(\theta)\Big)^2\right] = var_\theta(T(\boldsymbol{x})) + b_T^2(\theta)$$

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

Efficiency of estimators (2/5)

The MSE for the empirical mean is therefore

$$\begin{aligned}MSE = R(\overline{x}), \theta) &= var_\theta(\overline{x}) + b^2_{\overline{x}}(\theta) \\ &= \frac{var_\theta(\overline{x})}{n} + 0 \\ &= \frac{\sigma^2_\theta}{n}\end{aligned}$$

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

$$\frac{\sigma^2_\theta}{n} \quad \xrightarrow[n\to\infty]{} \quad 0$$

Efficiency of estimators (3/5)

The **Fisher information** for one observation is denoted $I(\theta)$. The Fisher information for the sample $x_1, x_2, ..., x_n$ 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

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

Efficiency of estimators (4/5)

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

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

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

Or equivalently, using the definition II

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

Efficiency of estimators (5/5)

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

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

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

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

We conclude that our estimator $\overline{x}$ is an efficient estimator for μ.

References

Bijma, F., Jonker M., Van der Vaart, A. (2016), An Introduction to Mathematical Statistics. Amsterdam University Press. ISBN 978 94 6298 5100

Hogg, R. V., McKean, J. W., Craig, A. T. (2019), Introduction to Mathematical Statistics, Eighth Edition. Pearson. ISBN 13: 978-0-13-468699-8

Course notes