

1 Point Estimation

1.1 Evaluating Estimators

$$\begin{aligned}
 \text{MSE}_{\hat{\Theta}}(\theta) &= \mathbb{E}_{\theta}(\hat{\Theta} - \theta)^2 \\
 &= \mathbb{E}_{\theta}(\hat{\Theta} - \mathbb{E}_{\theta}(\hat{\Theta}) + \mathbb{E}_{\theta}(\hat{\Theta}) - \theta)^2 \\
 &= \mathbb{E}_{\theta} \left(\hat{\Theta} - \mathbb{E}_{\theta}(\hat{\Theta}) \right)^2 + \left(\mathbb{E}_{\theta}(\hat{\Theta}) - \theta \right)^2 \\
 &= \mathbb{V}_{\theta}(\hat{\Theta}) + (B_{\hat{\Theta}}(\theta))^2
 \end{aligned}$$

So Mean Squared Error (MSE) is equal to the Variance of the estimator $\hat{\Theta}$ plus the squared bias of the estimator $\hat{\Theta}$.

Suppose $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} \text{Unif}[0, \theta]$, where $\theta > 0$ is the unknown parameter. Then find the following;

- $\hat{\Theta} = \max(X_1, X_2, \dots, X_n)$.
- Check for the unbiasedness and consistency of $\hat{\Theta}$.
- Determine the $\text{MSE}_{\hat{\Theta}}(\theta) \quad \forall \theta$.

We can easily see that the estimator is not unbiased because we are taking the maximum possible value of one particular sample, but $\max(X_1, \dots, X_n) \leq \theta$ then it's expected value will always fall short of θ .

More Precisely, we want to find

$$\mathbb{E}_{\theta}(\max(X_1, \dots, X_n))$$

Now,

$$\begin{aligned}
 F_{\hat{\Theta}}(x) &= \Pr_{\theta}(\max(X_1, \dots, X_n) \leq x) \\
 &= (\Pr_{\theta}(X_1 \leq x))^n \\
 &= \left(\frac{x}{\theta}\right)^n \\
 f_{\hat{\Theta}}(x) &= \frac{nx^{n-1}}{\theta^n}, \quad 0 \leq x \leq \theta
 \end{aligned}$$

So,

$$\begin{aligned}
 \mathbb{E}_{\theta}(\max(X_1, X_2, \dots, X_n)) &= \int_0^{\theta} x \cdot \frac{nx^{n-1}}{\theta^n} dx \\
 &= \left(\frac{n}{n+1}\right) \theta
 \end{aligned}$$

Which is clearly less than θ and therefore our estimator $\hat{\Theta}$ is not unbiased.

Now to check for Consistency we want to show that,

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\theta} (|\max (X_1, X_2, \dots, X_n) - \theta| > \epsilon) = 0$$

Now,

$$\begin{aligned} \mathbb{P}_{\theta} (|\max (X_1, X_2, \dots, X_n) - \theta| > \epsilon) &= \mathbb{P}_{\theta} (\theta - \max (X_1, X_2, \dots, X_n) > \epsilon) \\ &= \mathbb{P}_{\theta} (\max (X_1, X_2, \dots, X_n) < \theta - \epsilon) = \left(\frac{\theta - \epsilon}{\theta} \right)^n \end{aligned}$$

Then,

$$\lim_{n \rightarrow \infty} \left(\frac{\theta - \epsilon}{\theta} \right)^n = 0$$

and therefore are estimator $\hat{\Theta}$ is consistent.

Now to find the MSE we have to first find the variance of our estimator, i.e., $\mathbb{V}_{\theta}(\max (X_1, X_2, \dots, X_n))$ and then $(B_{\hat{\Theta}}(\theta))^2$;

$$\begin{aligned} \mathbb{V}_{\theta} &= \mathbb{E}_{\theta} [\hat{\Theta}^2] - \left(\mathbb{E}_{\theta} [\hat{\Theta}] \right)^2 \\ &= \int_0^{\theta} x^2 \cdot \frac{nx^{n-1}}{\theta^n} dx - \left[\int_0^{\theta} x \cdot \frac{nx^{n-1}}{\theta^n} dx \right]^2 \\ &= \frac{n\theta^2}{(n+1)^2(n+2)} \end{aligned}$$

and,

$$\begin{aligned} (B_{\hat{\Theta}}(\theta))^2 &= (\mathbb{E}_{\theta}[\hat{\Theta} - \theta])^2 \\ &= (\mathbb{E}_{\theta}[\hat{\Theta}] - \mathbb{E}_{\theta}[\theta])^2 \\ &= \frac{\theta^2}{(n+1)^2} \end{aligned}$$

Therefore,

$$\begin{aligned} \text{MSE}_{\hat{\Theta}}(\theta) &= \mathbb{V}_{\theta}(\hat{\Theta}) + (B_{\hat{\Theta}}(\theta))^2 \\ &= \frac{2\theta^2}{(n+1)(n+2)} \end{aligned}$$

When comparing two estimators such as $\hat{\Theta}_1$ and $\hat{\Theta}_2$, $\hat{\Theta}_1$ is better than $\hat{\Theta}_2$ if;

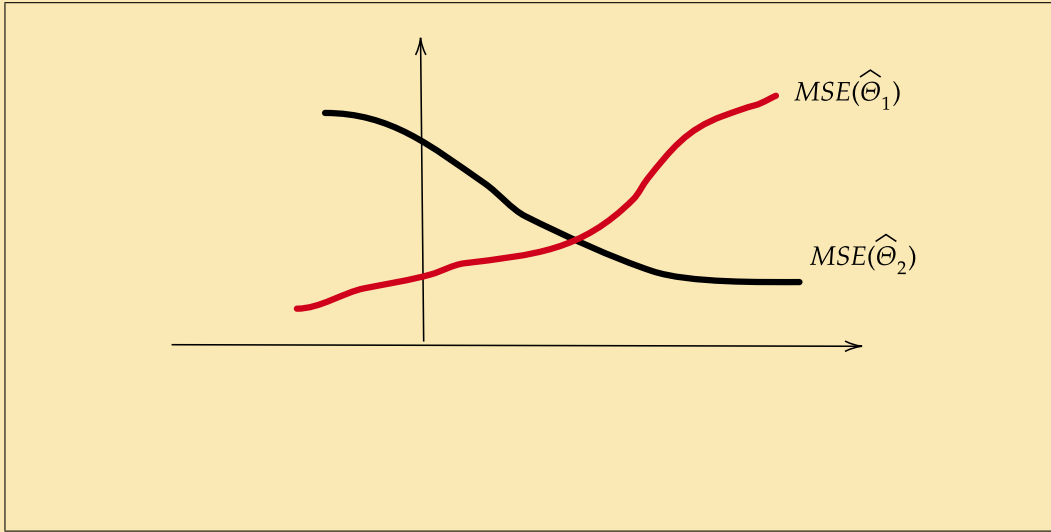


Figure 1: Comparing Mean Squared Errors

- $\text{MSE}_{\hat{\Theta}_1}(\theta) \leq \text{MSE}_{\hat{\Theta}_2}(\theta) \quad \forall \theta$
- $\text{MSE}_{\hat{\Theta}_1}(\theta) < \text{MSE}_{\hat{\Theta}_2}(\theta) \quad \text{for some } \theta$

Question: Suppose X_1, X_2, \dots, X_n are *iid* with $\mathbb{E}(X_i) = \mu (< \infty)$ and $\text{Var}(X_i) = \sigma^2 (< \infty)$, Compare the MSE of these estimators for μ

- $m_1 = X_1$
- $m_2 = \frac{X_1 + X_2}{2}$
- $m_n = \frac{X_1 + X_2 + \dots + X_n}{n}$

Note that all of them have the same expected values, i.e., $\mathbb{E}(m_1) = \mathbb{E}(m_2) = \dots = \mathbb{E}(m_n) = \mu$.

$$\text{MSE}(m_1) = \sigma^2$$

$$\text{MSE}(m_2) = \frac{\sigma^2}{2}$$

$$\text{MSE}(m_n) = \frac{\sigma^2}{n}$$

so the MSE grows smaller as n increases.

1.2 Maximum Likelihood Estimator

X_1, X_2, \dots, X_n are *iid* random variables from some population $f(x; \theta)$ where θ is the unknown parameter we want to estimate.

Take Observation X_1, X_2, \dots, X_n and find the joint density of them which will be the product of the marginals, we call this joint density a Likelihood function,

$$\mathcal{L}(\theta; x_1, x_2, \dots, x_n) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta)$$

then we can solve the following;

$$\max_{\theta} \mathcal{L}(\theta)$$

the solution to the above problem is $\hat{\theta}_{MLE}$.

Now,

Suppose X_1, X_2, \dots, X_n are *iid* Bern(p), then estimate p using MLE;

$$\mathcal{L}(p; x_1, \dots, x_n) = p^{\sum_{i=1}^n x_i} (1 - p)^{n - \sum_{i=1}^n x_i}$$

we want to maximise the above but we can take an increasing transformation and solve the following,

$$\begin{aligned} \max_p \ln(\mathcal{L}(p; x_1, \dots, x_n)) &= \left(\sum_{i=1}^n x_i \right) (\ln(p)) + \left(n - \sum_{i=1}^n x_i \right) (\ln(1 - p)) \\ \text{differentiating w.r.t } p; \quad &\frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1 - p} = 0 \\ \implies \sum_{i=1}^n x_i - p \sum_{i=1}^n x_i &= np - p \sum_{i=1}^n x_i \\ = \hat{p}_{MLE} &= \frac{\sum_{i=1}^n x_i}{n} \end{aligned}$$

So in this case it turns out to be just the sample mean.

Question: Suppose $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} \mathcal{N}(\mu, 1)$, where μ is unknown, find $\hat{\mu}_{MLE}$.

$$\begin{aligned}
 \mathcal{L}(\mu; x_1, x_2, \dots, x_n) &= f(x_1; \mu) f(x_2; \mu) \cdots f(x_n; \mu) \\
 \mathcal{L}(\mu) &= \prod_{i=1}^n f_{X_i}(x_i; \mu) = \left(\frac{1}{\sqrt{2\pi}} \right)^n \exp \left(-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2 \right) \\
 \ell(\mu) = \log \mathcal{L}(\mu) &= -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2 \\
 \frac{d\ell(\mu)}{d\mu} &= \frac{d}{d\mu} \left(-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2 \right) = \sum_{i=1}^n (x_i - \mu) = 0 \\
 \Rightarrow n\mu &= \sum_{i=1}^n x_i \\
 \mu_{MLE} &= \frac{1}{n} \sum_{i=1}^n X_i
 \end{aligned}$$

This is the sample mean, which makes intuitive sense as the best estimator of the population mean μ .