

Criteria of Estimation – Optimality

How should one choose an estimate when many are available? In other words, what criteria are to be used to determine the procedure of estimation? This prompts the question: how good is an estimate?

Suppose $X \sim P_\theta$ and $T(X)$ estimates $q(\theta)$. Then $|T(X) - q(\theta)|$ is the discrepancy in estimation. Does there exist $T(X)$ which can minimize this discrepancy (uniformly) for all θ ? No. Take $q(\theta) = \theta$. For $\theta = \theta_1$, $T(X) = \theta_1$ is the best, but this is not optimal for any other θ . Define $L(q(\theta), T(X))$ to be the loss due to estimating $q(\theta)$ by $T(X)$. Standard losses in the theory of estimation are

$$L(q(\theta), T(X)) = \begin{cases} |q(\theta) - T(X)| & \text{absolute error loss;} \\ (q(\theta) - T(X))^2 & \text{squared error loss.} \end{cases}$$

We have already noted that the loss cannot be minimized uniformly. Also, it depends on X which is random. Therefore, we average it over all samples. Then we get $R(q(\theta), T(X)) = E_\theta L(q(\theta), T(X))$ which is called the risk. Thus, we have $E_\theta(|T(X) - q(\theta)|) = \int |q(\theta) - T(x)| f(x|\theta) dx = \text{mean absolute error}$ and $E_\theta((T(X) - q(\theta))^2) = \int (q(\theta) - T(x))^2 f(x|\theta) dx = \text{mean square error (MSE)}$. Since these cannot be minimized uniformly for all θ , one may restrict $T(X)$ to some class of estimators and then choose the best in that class.

Unbiased estimators. $T(X)$ is said to be unbiased for $q(\theta)$ if $E_\theta(T(X)) = q(\theta)$ for all $\theta \in \Theta$.

Example. X_1, \dots, X_n i.i.d $\text{Exp}(\lambda)$, $q(\lambda) = 1/\lambda$. Then \bar{X} is an unbiased estimator of $q(\lambda)$ since $E(\bar{X}) = E(X) = 1/\lambda$ for all λ .

Note that MSE for unbiased estimators is just the variance of the estimate:

$$E_\theta(T(X) - q(\theta))^2 = E_\theta(T(X) - E_\theta(T(X)))^2 = \text{Var}(T(X)).$$

Unbiasedness means only that $E_\theta(T(X)) - q(\theta) = 0$. i.e., if used over and over again, on the average, underestimation will balance overestimation; no consideration is given to how often or by how much the estimate will depart from the parameter.

It is possible in many situations to find an estimate which is best among all unbiased estimates in terms of variance. Such an estimate is called *Uniformly Minimum Variance Unbiased Estimate* or UMVUE or *Best Unbiased Estimate*. Note that

- UMVUE may not exist;
- unbiasedness may be ridiculous;
- there may be better and simpler estimates which are not unbiased.

Example (Unbiased estimates do no exist). Suppose $X \sim \text{Binomial}(n, p)$. We want to estimate $q(p) = \frac{1}{p}$. Since $\hat{p} = X/n$ for both method of moments and MLE, n/X is the corresponding estimate, except when $X = 0$. (No estimate when $X = 0$.) Unbiasedness means,

$$E_p(T(X)) = \frac{1}{p} \text{ for all } p \in (0, 1).$$

(Note that $T(x)$ needs to be defined for all realizable x for computing $E(T)$.) Then, we must have,

$$\sum_{x=0}^n T(x) \binom{n}{x} p^x (1-p)^{n-x} = \frac{1}{p} \text{ for all } p \in (0, 1). \text{ i.e.,}$$

$$T(0)(1-p)^n + \sum_{x=1}^n T(x) \binom{n}{x} p^x (1-p)^{n-x} = \frac{1}{p} \text{ for all } p \in (0, 1).$$

As $p \rightarrow 0$, LHS $\rightarrow T(0)$ which is a real number, whereas RHS $\rightarrow \infty$. No such T exists.

Example (Unbiased estimates are silly). Suppose $X \sim \text{truncated Poisson}$:

$$P_\lambda(X = x) = \frac{\exp(-\lambda)\lambda^x/x!}{1 - \exp(-\lambda)}, x = 1, 2, \dots$$

Estimate $q(\lambda) = \exp(-\lambda)$, a positive quantity. Consider any unbiased estimate $T(X)$. Then,

$$\exp(-\lambda) = E(T(X)) = \sum_{x=1}^{\infty} T(x) \frac{\exp(-\lambda)\lambda^x/x!}{1 - \exp(-\lambda)},$$

so that

$$1 - \exp(-\lambda) = E(T(X)) = \sum_{x=1}^{\infty} T(x) \frac{\lambda^x}{x!}, \quad \forall \lambda > 0.$$

Now the power series expansion gives,

$$\begin{aligned} \text{LHS} &= 1 - \left[1 - \lambda + \frac{\lambda^2}{2!} - \frac{\lambda^3}{3!} + \dots \right] \\ &= - \sum_{x=1}^{\infty} \frac{(-1)^x \lambda^x}{x!} = \sum_{x=1}^{\infty} \frac{(-1)^{x+1} \lambda^x}{x!}. \end{aligned}$$

Therefore, we must have that

$$\sum_{x=1}^{\infty} \frac{(-1)^{x+1} \lambda^x}{x!} = \sum_{x=1}^{\infty} T(x) \frac{\lambda^x}{x!}, \quad \forall \lambda > 0.$$

Since two power series agree on an interval, their coefficients must be equal. Therefore the only unbiased estimate for $\exp(-\lambda)$ is

$$T(x) = \begin{cases} 1 & \text{if } x \text{ is odd;} \\ -1 & \text{if } x \text{ is even.} \end{cases}$$

i.e., our estimate $T(x) < 0$ if x is even!

Let $\hat{\theta}(x)$ be an estimator of θ . Consider $L(\theta, d) = (\theta - d)^2$, the squared error loss. Then

$$\begin{aligned} \text{MSE} &= R(\theta, d) = E_{\theta}(d(X) - \theta)^2 \\ &= E_{\theta} [d(X) - E_{\theta}(d(X)) + E_{\theta}(d(X)) - \theta]^2 \\ &= E_{\theta} [d(X) - E_{\theta}(d(X))]^2 + [E_{\theta}(d(X)) - \theta]^2 \\ &\quad + 2 [E_{\theta}(d(X)) - \theta] E_{\theta} [d(X) - E_{\theta}(d(X))] \\ &= \text{Var}_{\theta}(d(X)) + \text{Bias}^2(\theta), \end{aligned}$$

where $\text{Bias}(\theta) = E_{\theta}(d(X)) - \theta$. If $E_{\theta}(\hat{\theta}(X)) = \theta$ for all θ , i.e., $\hat{\theta}(X)$ is unbiased for θ , then $\text{MSE} = \text{Variance}$.

Example. Let X_1, \dots, X_n be i.i.d $N(\mu, \sigma^2)$. Consider the following estimates for σ^2 .

$$\begin{aligned} T_1(\mathbf{X}) &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{S^2}{n}, \\ T_2(\mathbf{X}) &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{S^2}{n-1}. \end{aligned}$$

Since $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 \sim \sigma^2 \chi_{n-1}^2$, $E(T_1) = \frac{n-1}{n} \sigma^2$ and $E(T_2) = \sigma^2$. Thus,

T_2 is unbiased and T_1 is not. However, compare their MSE.

$$\begin{aligned}
\text{MSE}(T_1) &= E(T_1 - \sigma^2)^2 = \text{Var}(T_1) + \text{Bias}^2 \\
&= \text{Var}\left(\frac{S^2}{n}\right) + \left(E\left(\frac{S^2}{n}\right) - \sigma^2\right)^2 \\
&= \frac{1}{n^2} \text{Var}(S^2) + \left(\frac{n-1}{n} \sigma^2 - \sigma^2\right)^2 \\
&= \frac{1}{n^2} 2(n-1)\sigma^4 + \frac{1}{n^2} \sigma^4 = \sigma^4 \left[\frac{2(n-1)}{n^2} + \frac{1}{n^2} \right] \\
&= \sigma^4 \left(\frac{2n-1}{n^2} \right), \text{ and} \\
\text{MSE}(T_2) &= \text{Var}(T_2) = \text{Var}\left(\frac{S^2}{n-1}\right) \\
&= \frac{1}{(n-1)^2} 2(n-1)\sigma^4 = \sigma^4 \left(\frac{2}{n-1} \right).
\end{aligned}$$

Note that,

$$\begin{aligned}
\frac{2}{n-1} - \frac{2n-1}{n^2} &= \frac{2n^2 - (2n-1)(n-1)}{n^2(n-1)} \\
&= \frac{2n^2 - \{2n^2 - 2n - n + 1\}}{n^2(n-1)} = \frac{3n-1}{n^2(n-1)} > 0.
\end{aligned}$$

Thus, T_1 has a smaller MSE than T_2 for all σ^2 even though it is not unbiased. T_2 is often preferred because T_1 can underestimate σ^2 which is undesirable.