## △ Evaluation Methods on the quality of an estimator $\hat{\theta}$

Many different ways to estimate a parameter $\theta$

$$\hat{\theta} = \hat{\theta}(\vec{X})$$

To tell the "quality" of $\hat{\theta}$ an estimator, $\hat{\theta}$, there are various criteria

① **Bias** — measures how far expect $\hat{\theta}$ is from $\theta$ on average

$$\text{Bias}(\hat{\theta}) := E(\hat{\theta}) - \theta_0$$

The estimator $\hat{\theta}$, for any value of $\theta$,

↳ is **unbiased** if the bias $(\hat{\theta}) = 0$

↳ **overestimates** $\theta$ if the bias $> 0$.

↳ **underestimates** $\theta$ if the bias $< 0$.

If the bias is nonzero for finite sample size $n$, but $\to 0$ as $n \to \infty$, then it is **asymptotically unbiased**.

**E.g.** A random sample, i.i.d., $\{X_1, X_2, \dots, X_n\}$ from a distribution with mean $\mu$ and variance $\sigma^2$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$S = \sqrt{S^2}$$

**E.g.2** (Uniform) $X_i \sim \text{Uniform}(0, \theta)$

$$\hat{\theta}_{max} = 2\bar{X}_n \quad , \qquad E(\hat{\theta}_{max}) = 2E(\bar{X}_n) =$$

$$\hat{\theta}_{MLE} = X_{(n)} \quad , \qquad E(\hat{\theta}_{MLE}) = \frac{n}{n+1}\theta_0$$

To compute $E(X_{(n)})$, find cdf of $X_{(n)}$, then $\frac{d}{dx}$ to find pdf.

② **Mean Squared Error (MSE)** measures the variability of $\hat{\theta}$

$$MSE(\hat{\theta}) := E\left[(\hat{\theta} - \theta_0)^2\right]$$

$$\begin{aligned} MSE(\hat{\theta}) &= E\left[(\hat{\theta} - \theta_0)^2\right] = E\left[(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta_0)^2\right] \\ &= E\left[(\hat{\theta} - E(\hat{\theta}))^2\right] + E\left[(E(\hat{\theta}) - \theta_0)^2\right] + 2E\left[(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta_0)\right] \\ &= Var(\hat{\theta}) + \left(E(\hat{\theta}) - \theta_0\right)^2 + 2\left(E(\hat{\theta}) - \theta_0\right)\left[E(\hat{\theta}) - E(\hat{\theta})\right] \\ &= Var(\hat{\theta}) + (\text{Bias})^2 \end{aligned}$$

A lower MSE is generally desired. For an unbiased $\hat{\theta}$, $MSE(\hat{\theta}) = Var(\hat{\theta})$.

③ (Relative) Efficiency.

**Def 1:** Given two estimators, $\hat{\theta}_1$ and $\hat{\theta}_2$, of a parameter $\theta$.

$$\text{Eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{MSE(\hat{\theta}_2)}{MSE(\hat{\theta}_1)}$$

**Def 2:** Given two unbiased estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ of $\theta$.

$$\text{Eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)}$$

The ratio is called the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$.

E.g. $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$ ⟺ $\text{eff}(\hat{\theta}_1, \hat{\theta}_2) > 1$

**Report:** If an unbiased estimator that has the smallest variance among all other unbiased ones, for all $\theta$, we call such $\hat{\theta}$ as the uniformly minimum variance unbiased estimator, or UMVUE

④ Consistency

An estimator $\hat{\theta}_n$ is a consistent estimator of $\theta$ if

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1 \quad , \quad \forall\ \varepsilon > 0$$

denoted by $\hat{\theta}_n \xrightarrow{P} \theta$ as $n \to \infty$

## △ Cramer-Rao Inequality

**Thm.** Let $X_1, \dots, X_n$ be i.i.d from $\pi(X|\theta)$. Let $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$

be any unbiased estimator of a parameter $\theta$ whose true value is $\theta_0$.

Then, under conditions assumption on $\pi(X|\theta)$,

$$MSE(\hat{\theta}) \overset{\text{def}}{=} Var(\hat{\theta}) \geq \boxed{\frac{1}{n\mathcal{I}(\theta)}}$$

Cramer-Rao's lower bound

**Idea of proof:** Let $Z = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log \pi(X|\theta)$

$$= \sum_{i=1}^{n} \frac{\frac{\partial}{\partial\theta}\pi(X|\theta)}{\pi(X|\theta)}$$

with $E(Z) = 0$ (Claim before) and $Var(Z) = n\mathcal{I}(\theta)$ (Claim before)

Then consider the correlation coefficient of $Z$ and $\hat{\theta}$ is less than 1 in absolute value

$$\underbrace{Cor^2(Z, \hat{\theta})}_{\text{Since } \leq 1} \leq \frac{Var(Z) \cdot Var(\hat{\theta})}{n\mathcal{I}(\theta)}$$

To show $Cor(Z, \hat{\theta}) = 1$,

$$\begin{aligned} Cov(Z, \hat{\theta}) &= E(Z\hat{\theta}) - E(Z)E(\hat{\theta}) \\ &= E(Z\hat{\theta}) \\ &= \int \cdots \int \hat{\theta} \left[\sum_{i=1}^{n} \frac{\frac{\partial}{\partial\theta}\pi(X|\theta)}{\pi(X|\theta)}\right] \prod_{i=1}^{n} \pi(X|\theta)\, dx_1 \cdots dx_n \\ &= \cdots = \frac{\partial}{\partial\theta} E(\hat{\theta}) = \frac{\partial}{\partial\theta}(\theta) = 1 \end{aligned}$$

cont. on next page

Since unbiased

**E.g.2.4** (Poisson)

A random sample $X_1, \dots, X_n \sim Poi(\lambda)$

$$f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad , \quad x = 0, 1, 2, \dots$$

① Find the MLE of $\lambda$.

② Show that $\hat{\lambda}_{MLE}$ is efficient

Soln. ① $L(\vec{X}|\lambda) = \prod_{i=1}^{n} f(x_i|\lambda)$

② log-likelihood function

$$\ell(\lambda) = \sum_{i=1}^{n} \log f(x_i|\lambda)$$

③ $\frac{\partial}{\partial\lambda}\ell(\lambda) = \sum_{i=1}^{n} \frac{\partial}{\partial\lambda} \log f(x_i|\lambda)$

$$= \sum_{i=1}^{n} \frac{x_i}{\lambda} - n$$

$$\frac{\partial}{\partial\lambda}\ell(\lambda) = 0 \quad \to \quad \sum_{i=1}^{n} \frac{x_i}{\lambda} = n \Rightarrow \hat{\lambda} = \frac{\sum x_i}{n} = \bar{X}$$

$$\frac{\partial}{\partial\lambda}\log f(x|\lambda) = \frac{\frac{\partial}{\partial\lambda} f(x|\lambda)}{f(x|\lambda)} = \frac{\partial}{\partial\lambda}\left[\log \lambda^x + \log e^{-\lambda} - \log x!\right]$$

$$= \frac{\partial}{\partial\lambda}\left[x \log \lambda - \lambda - \log x!\right]$$

$$= \frac{x}{\lambda} - 1$$

② $\hat{\lambda} = \bar{X}$ , $E(\hat{\lambda}) = E(\bar{X}) = \frac{1}{n}\sum E(X_i) = \frac{1}{n}\sum \lambda$

$$Var(\hat{\lambda}) = Var(\bar{X}) = \frac{1}{n^2} Var\left(\sum X_i\right) \overset{i.i.d}{=} \frac{1}{n^2}\sum Var(X)$$

$$= \frac{1}{n^2} \cdot n \cdot Var(X) = \frac{\lambda}{n}$$

Verify it is the lower bound

$$\mathcal{I}(\lambda) = -E\left[\frac{\partial^2}{\partial\lambda^2}\left[\log f(x|\lambda)\right]\right] = E\left[\frac{\partial}{\partial\lambda}\left[\frac{x}{\lambda} - 1\right]\right]$$

$$= -E\left[-\frac{x}{\lambda^2}\right] = \frac{1}{\lambda^2}E(x) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

Since $\frac{\partial}{\partial\lambda}\log f(x|\lambda) = \frac{\partial}{\partial\lambda}\log\left(\frac{\lambda^x e^{-\lambda}}{x!}\right) = \frac{\partial}{\partial\lambda}\left[x\log\lambda - \lambda - \log x!\right]$

$$= \frac{x}{\lambda} - 1$$

$$\frac{\partial^2}{\partial\lambda^2}\log f(x|\lambda) = -\frac{x}{\lambda^2}$$

The Cramer-Rao lower bound:

$$\frac{1}{n\mathcal{I}(\lambda)} = \frac{1}{n/\lambda} = \frac{\lambda}{n}$$

Since $Var(\hat{\lambda}_{MLE}) = \frac{\lambda}{n}$, then $\hat{\lambda}_{MLE}$ is efficient.

**Example** (cont.) (Uniform)

Let $X_1, \dots, X_n \sim U(0, \theta)$

Recall that $\hat{\theta}_{MOM} = 2\bar{X}_n$ and $\hat{\theta}_{MLE} = X_{(n)}$

Because: $E(\hat{\theta}_{MOM}) = E(2\bar{X}) = 2E(\bar{X})$

$$= 2E\left(\frac{1}{n}\sum X_i\right)$$

$$= \frac{2}{n}\sum E(X_i) = \frac{2}{n} \cdot n \cdot \frac{\theta}{2} = \theta \quad , \quad \forall\ \theta$$

As for $\hat{\theta}_{MLE}$,

$$\begin{aligned} E(\hat{\theta}_{MLE}) &= E(X_{(n)}) \\ &= \int_0^\theta \frac{n x^{n-1}}{\theta^n} x\, dx = n\int_0^\theta \frac{x^n}{\theta^n}\, dx \\ &= \frac{n}{\theta^n}\left[\frac{1}{n+1} x^{n+1}\right]_0^\theta = \frac{n}{n+1}\theta \end{aligned}$$

$\to$ biased, underestimate $\theta$, but asymptotically unbiased.

$P(X_i < x) = \int_0^x \frac{1}{\theta}\, dx = \frac{x}{\theta}$

As for distribution of $X_{(n)}$: Find cdf $\to$ take derivative on cdf

$$\begin{aligned} P(X_{(n)} \geq x) &= 1 - P(X_{(n)} < x) \\ &= 1 - P(X_1 < x, X_2 < x, \dots, X_n < x) \\ &= 1 - P(X_1 < x)P(X_2 < x) \cdots P(X_n < x) \\ &= 1 - \left(\frac{x}{\theta}\right)^n \end{aligned}$$

cdf: $F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = \left(\frac{x}{\theta}\right)^n \quad , \quad 0 \leq x \leq \theta$

pdf: $\frac{d}{dx}F_{X_{(n)}}(x) = \frac{d}{dx}\frac{x^n}{\theta^n} = \frac{n x^{n-1}}{\theta^n} \quad , \quad 0 \leq x \leq \theta$

Remark: We can scale an unbiased estimator from $\hat{\theta}_{MLE}$.

by letting $\hat{\theta}_1 = \frac{n+1}{n}\hat{\theta}_{MLE} = \frac{n+1}{n}X_{(n)}$

**Mean-Square Error:** $X \sim U(0, \theta)$ ; $Var(X) = \frac{\theta^2}{12}$ ; i.i.d. sample $E(\bar{X}) = \mu$ , $Var(\bar{X}) = \frac{\sigma^2}{n}$

$$MSE(\hat{\theta}_{MOM}) \overset{\text{unbiased}}{=} Var(\hat{\theta}_{MOM}) = Var(2\bar{X}) = 4Var(\bar{X})$$

$$= 4\frac{Var(X_i)}{n} = \frac{4}{n}\frac{\theta^2}{12} = \frac{\theta^2}{3n}$$

$$\begin{aligned} MSE(\hat{\theta}_{MLE}) &= Var(\hat{\theta}_{MLE}) + (\text{Bias})^2 \\ &= Var(\hat{\theta}_{MLE}) + \left(\frac{n}{n+1}\theta - \theta\right)^2 \\ &= \frac{1}{(n+1)^2(n+2)} n\theta^2 + \frac{1}{(n+1)^2}\theta^2 = \frac{\theta^2}{(n+1)^2}\left[\frac{n}{n+2} + 1\right] = \frac{2\theta^2}{(n+1)(n+2)} \end{aligned}$$

As for $Var(\hat{\theta}_{MLE})$:

$$\begin{aligned} Var(\hat{\theta}_{MLE}) &= E(\hat{\theta}_{MLE}^2) - E(\hat{\theta}_{MLE})^2 = E(X_{(n)}^2) - E(X_{(n)})^2 \\ &= \int_0^\theta x^2 \frac{n x^{n-1}}{\theta^n} dx - \left[\frac{n}{(n+1)}\theta\right]^2 \\ &= \frac{n}{\theta^n}\int_0^\theta x^{n+1} dx - \frac{n^2}{(n+1)^2}\theta^2 \\ &= \frac{n}{(n+2)}\frac{1}{\theta^n}\theta^{n+2} - \frac{n^2}{(n+1)^2}\theta^2 \\ &= n\theta^2\left[\frac{1}{(n+2)} - \frac{n}{(n+1)^2}\right] = \frac{n^2+2n+1-n^2-2n}{(n+2)(n+1)^2} \cdot n\theta^2 \\ &= \frac{1}{(\quad)^2} \cdot n\theta^2 \end{aligned}$$

$$\text{Eff}(\hat{\theta}_{MLE}, \hat{\theta}_{MOM}) = \frac{MSE(\hat{\theta}_{MOM})}{MSE(\hat{\theta}_{MLE})} = \frac{n^2+2n+2}{6n} > 1 \quad \text{for } n \geq 3$$

So $\hat{\theta}_{MLE}$ is more efficiency compared to $\hat{\theta}_{MOM}$.