Cramer-Rao lower bound

Recap: MSE

Let $\hat{\theta}$ be an estimator of θ. The MSE of $\hat{\theta}$ is

$$\mathbb{E}_\theta\left[(\hat{\theta}-\theta)^2\right]$$

$$= \mathrm{Var}_\theta(\hat{\theta}) + \left(\mathrm{Bias}(\hat{\theta})\right)^2$$

MSE and consistency

<u>Def</u>: If $\hat{\theta} \xrightarrow{P} \theta$, then we say $\hat{\theta}$ is a <u>consistent</u> estimator of θ

<u>Theorem</u>: If $MSE(\hat{\theta}) \to 0$ as $n \to \infty$, then $\hat{\theta} \xrightarrow{P} \theta$

<u>Pf</u>: WTS $\forall \varepsilon > 0$, $P(|\hat{\theta} - \theta| > \varepsilon) \to 0$

Let $\varepsilon > 0$.

$$P(|\hat{\theta} - \theta| > \varepsilon) = P(|\hat{\theta} - \theta|^2 > \varepsilon^2)$$

$$\leq \frac{E[(\hat{\theta} - \theta)^2]}{\varepsilon^2} \quad \text{(Markov's)}$$

$$= \frac{MSE(\hat{\theta})}{\varepsilon^2} \to 0 \quad //$$

(i.e., if $Bias(\hat{\theta}) \to 0$ and $Var(\hat{\theta}) \to 0$, then $\hat{\theta}$ is consistent for θ)

Best unbiased estimators

So suppose we restrict ourselves to unbiased estimators
($\text{Bias}(\hat{\theta}) = 0$)

Def: $\hat{\theta}$ is a best unbiased estimator of θ if
$MSE(\hat{\theta}) \leq MSE(\hat{\theta}^*)$ for all other unbiased
estimators $\hat{\theta}^*$ (i.e., if $Var(\hat{\theta}) \leq Var(\hat{\theta}^*)$)

Goal could be to find a best unbiased estimator

Harald Cramér   CR Rao   (still alive! age 102)

Cramér-Rao lower bound

<u>Thm</u>: Let $X_1, \dots, X_n$ be a sample from a distribution with probability function $f(x|\theta)$, and let $\hat{\theta}$ be an unbiased estimator of $\theta \in \mathbb{R}$. Then, under regularity conditions,

$$\left.\operatorname{Var}(\hat{\theta}) \geq \frac{1}{\mathcal{I}(\theta)}\right\} \text{ Cramér-Rao lower bound (CRLB)}$$

<u>Ex</u>: $X_1, \dots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$

MLE: $\hat{\lambda} = \overline{X}$

$$\operatorname{Var}(\hat{\lambda}) = \operatorname{Var}\left(\tfrac{1}{n} \textstyle\sum_i X_i\right) = \frac{1}{n^2} \textstyle\sum_i \operatorname{Var}(X_i) = \frac{\lambda}{n}$$

$$\mathcal{I}(\lambda) = \frac{n}{\lambda} \Rightarrow \text{CRLB} = \frac{\lambda}{n}$$

$$E[\hat{\lambda}] = \lambda \Rightarrow \text{Bias}(\hat{\lambda}) = 0, \operatorname{Var}(\hat{\lambda}) = \text{CRLB}$$

$\Rightarrow \hat{\lambda} = \overline{X}$ is a <u>best unbiased estimator</u> of λ

Pf: Recall Cauchy-Schwarz inequality:

$$[Cov(X,Y)]^2 \le Var(X)\,Var(Y)$$

$$\Rightarrow Var(X) \ge \frac{[Cov(X,Y)]^2}{Var(Y)}$$

Here, we apply C-S to $\hat{\theta}$ and $u(\theta)$ $\left(u(\theta) = \frac{\partial}{\partial\theta} \log f(X_1,\dots,X_n|\theta)\right)$

$$Cov(\hat{\theta}, u(\theta)) = \mathbb{E}[\hat{\theta}u(\theta)] - \mathbb{E}[\hat{\theta}]\underbrace{\mathbb{E}[u(\theta)]}_{0}$$ (under regularity)

$$= \mathbb{E}[\hat{\theta}u(\theta)]$$

$$= \mathbb{E}\left[\hat{\theta}\,\frac{\partial}{\partial\theta}\log f(X_1,\dots,X_n|\theta)\right] = \mathbb{E}\left[\hat{\theta}\,\frac{\frac{\partial}{\partial\theta}f(X_1,\dots,X_n|\theta)}{f(X_1,\dots,X_n|\theta)}\right]$$

$$= \int \hat{\theta}(x_1,\dots,x_n)\left[\frac{\partial}{\partial\theta}f(x_1,\dots,x_n|\theta)\right]dx_1,\dots,dx_n$$

(regularity conditions)

$$= \frac{\partial}{\partial\theta}\int \hat{\theta}(x_1,\dots,x_n)f(x_1,\dots,x_n|\theta)\,dx_1\dots dx_n = \frac{\partial}{\partial\theta}\mathbb{E}[\hat{\theta}]$$

$$\Rightarrow \text{C-S}: \quad Var(\hat{\theta}) \ge \frac{\left(\frac{\partial}{\partial\theta}\mathbb{E}[\hat{\theta}]\right)^2}{Var(u(\theta))} = \frac{\left(\frac{\partial}{\partial\theta}\mathbb{E}[\hat{\theta}]\right)^2}{\mathcal{I}(\theta)}$$

If $\hat{\theta}$ is unbiased, $\mathbb{E}[\hat{\theta}] = \theta \Rightarrow \frac{\partial}{\partial\theta}\mathbb{E}[\hat{\theta}] = 1 \Rightarrow Var(\hat{\theta}) \ge \frac{1}{\mathcal{I}(\theta)}$ //

Recall : under certain conditions, the MLE $\hat{\theta}$ of $\theta \in \mathbb{R}$ has the following properties:

① $\hat{\theta} \xrightarrow{P} \theta$ (Consistent)

② $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \mathcal{I}_1^{-1}(\theta))$

(iid data)

$\mathcal{I}(\theta) = n\,\mathcal{I}_1(\theta)$

$\Rightarrow \hat{\theta} \approx N(\theta, \mathcal{I}^{-1}(\theta))$ (asymptotic MLE)

asymptotically unbiased ($\text{Bias}(\hat{\theta}) \to 0$)

$\text{CRLB} = \mathcal{I}^{-1}(\theta)$

So asymptotically, $\text{Var}(\hat{\theta}) \approx \text{CRLB}$