Lecture 32: Variance and unbiased estimators

Ciaran Evans

Best unbiased estimators

Suppose we restrict ourselves to **unbiased** estimators.

Definition (best unbiased estimator): $\hat{\theta}$ is a best unbiased estimator of θ if

$$\mathrm{Var}(\hat{\theta}) \leq \mathrm{Var}(\hat{\theta}^*)$$

for all other unbiased estimators $\hat{\theta}^*$

(a.k.a. "minimum variance unbiased estimator" (MVUE))

Cramér-Rao lower bound

A is PSD if $v^T A v \geq 0 \quad \forall v$

Theorem: Let $X_1, \dots, X_n$ be an iid sample from a distribution with probability function $f(x|\theta)$, and let $\hat{\theta}$ be an unbiased estimator of $\theta \in \mathbb{R}$ calculated from $X_1, \dots, X_n$.

Under regularity conditions,

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{\mathcal{I}(\theta)}$$ ← Cramér-Rao lower bound (CRLB)

$\mathcal{I}(\theta) = \mathrm{Var}(\ell'(\theta))$ $\qquad \ell(\theta) = \log L(\theta|X)$

$\overset{\text{(under regularity conditions)}}{=} -\mathbb{E}[\ell''(\theta)]$

Generalization to vectors $\theta \in \mathbb{R}^d$:

$$\mathrm{Var}(\hat{\theta}) \succeq \mathcal{I}^{-1}(\theta)$$

where $A \succeq B$ means $A - B$ is PSD

Example

Suppose $X_1, \ldots, X_n \overset{iid}{\sim} Poisson(\lambda)$

$$L(\lambda | x) = \prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}$$

$$\Rightarrow \quad \ell(\lambda) = \sum_{i=1}^{n} \left[x_i \log(\lambda) - \lambda - \log(x_i!) \right]$$

$$\ell'(\lambda) = \sum_{i=1}^{n} \left[\frac{x_i}{\lambda} - 1 \right]$$

$$\mathcal{I}(\lambda) = Var(\ell'(\lambda)) = \sum_{i=1}^{n} Var\left(\frac{x_i}{\lambda} - 1 \right) = \sum_{i=1}^{n} \frac{1}{\lambda^2} Var(X_i)$$

$$= \frac{1}{\lambda^2} \cdot n\lambda = \frac{n}{\lambda}$$

MLE : $\hat{\lambda} = \bar{X}$ $\qquad Var(\bar{X}) = \dfrac{\lambda}{n} = \mathcal{I}^{-1}(\lambda)$

$\mathbb{E}[\hat{\lambda}] = \mathbb{E}[X] = \lambda$ $\qquad \Rightarrow$ MLE $\hat{\lambda}$ attains CRLB

unbiased!

$\Rightarrow$ MLE $\hat{\lambda}$ is best unbiased estimator of λ

logistic regression model:

$$Y_i \sim \text{Bernoulli}(p_i)$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta^T X_i$$

$$\frac{\partial l}{\partial \beta} = X^T(Y - p) \qquad p = \begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix}$$

$$\mathcal{I}(\beta) = X^T W X \qquad W = \begin{bmatrix} p_1(1-p_1) & & \\ & p_2(1-p_2) & \\ & & \ddots \end{bmatrix}$$

$\Rightarrow$ CRLB for an <u>unbiased estimator</u> of β is $\mathcal{I}^{-1}(\beta) = (X^T W X)^{-1}$

Also: $\hat{\beta} \approx N(\beta, \mathcal{I}^{-1}(\beta))$

$\Rightarrow$ asymptotically, $Var(\hat{\beta}) \approx$ CRLB

Example

Suppose that $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. want: CRLB for σ^2

$$L(\mu, \sigma^2 \mid X) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\right\}$$

$$\ell(\mu, \sigma^2 \mid X) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(X_i-\mu)^2$$

$$I(\sigma^2) = Var\left(\frac{\partial \ell}{\partial \sigma^2}\right) = Var\left(\frac{1}{2\sigma^4}\sum_{i=1}^{n}(X_i-\mu)^2\right)$$

$$= Var\left(\frac{1}{2\sigma^2}\underbrace{\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2}_{\chi^2_n}\right)$$

$$= \left(\frac{1}{2\sigma^2}\right)^2 Var(\chi^2_n)$$

$$= \left(\frac{1}{2\sigma^2}\right)^2 \cdot 2n = \frac{n}{2\sigma^4}$$

Example

$s^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2 \qquad \mathbb{E}[s^2] = \sigma^2$

Suppose that $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.

$\mathcal{I}(\sigma^2) = \frac{n}{2\sigma^4} \qquad \Rightarrow \text{CRLB}: \frac{2\sigma^4}{n}$

$\text{Var}(s^2) = \frac{2\sigma^4}{n-1} > \frac{2\sigma^4}{n} \quad \Rightarrow s^2$ does not attain CRLB

In fact, we can't attain CRLB unless we know μ

If we know μ: best unbiased estimator is $\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$

$\mathbb{E}[\frac{1}{n}\sum_i (X_i-\mu)^2] = \sigma^2$ ✓

$\text{Var}(\frac{1}{n}\sum_i (X_i-\mu)^2) = \frac{2\sigma^4}{n}$ ✓ (CRLB)

Example

Suppose that $X_1, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.

CRLB : $\frac{2\sigma^4}{n}$

MLE: $\hat{\sigma}^2 = \frac{1}{n} \sum_i (X_i - \bar{X})^2$ (if μ is unknown)

$Var(\hat{\sigma}^2) = \frac{2\sigma^4(n-1)}{n^2}$

$Bias(\hat{\sigma}^2) = -\frac{\sigma^2}{n}$

$Bias(\hat{\sigma}^2) \to 0$ as $n \to \infty$ (asymptotically unbiased)

$\frac{Var(\hat{\sigma}^2)}{CRLB} = \frac{(n-1)}{n} \to 1$ as $n \to \infty$

Why MLEs are nice

Let θ be a parameter of interest, and $\widehat{\theta}$ be the maximum likelihood estimator from a sample of size n. Under regularity conditions, $\widehat{\theta}$ satisfies the following properties:

- $\widehat{\theta} \xrightarrow{p} \theta$ (consistency)
- $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, \underbrace{\mathcal{I}_1^{-1}(\theta)}_{\text{variance}})$ (asymptotic normality)

$$\Rightarrow \quad \hat{\theta} \approx N(\theta, \underbrace{\mathcal{I}^{-1}(\theta)}_{\text{CRLB}}) \qquad \mathcal{I}(\theta) = n\mathcal{I}_1(\theta)$$

<u>Asymptotic efficiency</u>: asymptotically,

$$\frac{Var(\hat{\theta})}{CRLB} \to 1$$

Sufficient statistics

Question: Given an unbiased estimator, can I improve its variance?

- Answering this requires us to introduce a new concept: sufficient statistics

Definition (sufficient statistic): Let $X_1, \dots, X_n$ be a sample from a distribution $f(x|\theta)$. Let $T \doteq T(X_1, \dots, X_n)$ be a statistic. T is a <u>sufficient statistic</u> for θ if the conditional distribution of $X_1, \dots, X_n | T$ does not depend on θ

Example

Suppose $X_1, ..., X_n \overset{iid}{\sim} Poisson(\lambda)$

Example

Suppose $X_1, ..., X_n \overset{iid}{\sim} Bernoulli(p)$