

## Evaluation of the estimators

- need 2 criterion to evaluate the estimation
- the most used is the Mean Squared Error (MSE)

Definition 7.3.1 The Mean Squared Error of an estimator  $W$  of  $\theta$  is the function of  $\theta$  defined as  $E_{\theta}[(W - \theta)^2]$

In general any function that increases as  $|W - \theta|$  increases could work  
We use the MSE:  $\rightarrow$  because it's easy to treat mathematically

$$\rightarrow E_{\theta}[(W - \theta)^2] = \underbrace{\text{Var}_{\theta}(W)}_{\text{variance}} + \underbrace{\left(E_{\theta}[W - \theta]\right)^2}_{\text{squared bias}}$$

Definition 7.3.2: the bias of a point estimator  $W$  is the difference between its expected value and  $\theta$ . If  $W$  has bias 0, then it is called unbiased

Example

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \leftarrow \text{MLE } \sigma^2$$

$$E[\hat{\sigma}^2] = \frac{1}{n} E\left[\sum x_i^2 - 2\bar{x} \sum x_i + n\bar{x}^2\right] = \frac{n}{n} \left(\sigma^2 - \frac{2\sigma^2}{n} + \frac{\sigma^2}{n}\right) = \frac{n-1}{n} \sigma^2$$

$$\text{bias}(\hat{\sigma}^2) = E\left[\frac{n-1}{n} \sigma^2 - \sigma^2\right] = -\frac{\sigma^2}{n} \quad \leftarrow \text{biased estimator}$$

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

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

$$\text{MSE}(\hat{\sigma}^2) < \text{MSE}(S^2)$$

$$\text{note the } \text{MSE}(S^2) = \text{Var}(S^2) = \frac{2}{n-1} \sigma^4$$

(bias-variance trade-off): we can accept some bias to reduce the variance and hence the MSE

Note: • MSE is a function of  $\theta$ , so it does not necessarily be the case that  $\exists$  an estimator whose MSE is uniformly the smallest  $\forall \theta \in \Theta$

e.g.:  $\hat{\theta} = 17$ ,  $\hat{\theta}$  is perfect in  $\theta = 17$ , it is terrible everywhere else

- if an estimator is unbiased,  $\text{MSE} = \text{var}$
- if we restrict ourselves to unbiased estimators, we need to find that with uniformly smallest variance

Definition 7.3.7. An estimator  $W^*$  is a best unbiased estimator for  $\tau(\theta)$  if it satisfies

$$E_{\theta}[W^*] = \tau(\theta) \quad \forall \theta \in \Theta, \text{ and, } \forall \text{ other estimator with } E_{\theta}[W] = \tau(\theta), \text{ we have}$$

$$\text{Var}_{\theta}[W^*] \leq \text{Var}_{\theta}[W] \quad \forall \theta \in \Theta$$

$W^*$  is also called Uniform Minimum Variance Unbiased Estimator of  $\tau(\theta)$

UMVUE

Example  $X_1, \dots, X_n$  iid from  $X_j \sim \text{Poisson}(\lambda)$

$\bar{X}$  sample mean,  $S^2$  sample variance, both unbiased  $E(\bar{X}) = \lambda$ ,  $E(S^2) = \lambda$

• how can we decide which is the best?

- not always easy to find  $\text{Var}(\hat{\theta})$  for all estimators, e.g. here  $\text{Var}(S^2)$
- ' ' ' to check which one has the minimum variance in the whole  $\Theta$
- there may exist many other unbiased estimators, e.g.  $W_n(\bar{X}, S^2) = a\bar{X} + (1-a)S^2$  are good candidates

Since we cannot check the variance of all unbiased estimator, we proceed by

- finding a lower bound for the variance of ALL unbiased estimators of  $\theta$ ;
- if  $\text{Var}_\theta(W^*) = B(\theta)$ , then  $W^*$  is UMVUE;

Theorem 7.3.9 (Cramer - Rao inequality). Let  $X_1, \dots, X_n$  be a sample with pdf  $f(x; \theta)$  and let  $W(x) = W(X_1, \dots, X_n)$  be any estimator s.t.

*we can change the order of derivation and integration then*  $\frac{d}{d\theta} E_\theta[W(\theta)] = \int_X \frac{\partial}{\partial \theta} [W(x) f(x; \theta)] dx$  and  $\text{Var}(W(x)) < \infty$

*it works for pdf, using summation instead of integration*

$$\text{Var}_\theta(W(x)) \geq \frac{\left( \frac{d}{d\theta} E_\theta[W(x)] \right)^2}{E_\theta \left[ \left( \frac{\partial}{\partial \theta} \log f(x; \theta) \right)^2 \right]}$$

Proof Simply apply the Cauchy-Schwarz Inequality ( $\text{Cov}(X, Y)^2 \leq \text{Var}(X) \text{Var}(Y)$ ) with  $X = W(x)$  and  $Y = \frac{\partial}{\partial \theta} \log f(x; \theta)$

$$\begin{aligned} \frac{d}{d\theta} E_\theta[W(x)] &= \int_X W(x) \frac{\partial}{\partial \theta} f(x; \theta) dx \cdot \frac{f(x; \theta)}{f(x; \theta)} \\ &= E_\theta \left[ W(x) \frac{\partial}{\partial \theta} \log f(x; \theta) \right] \\ &= E_\theta \left[ W(x) \frac{\partial}{\partial \theta} \log f(x; \theta) \right] \end{aligned}$$

*$E[XY]$  one piece of the  $\text{Cov}(X, Y)$*

$$\begin{aligned} E \left[ \frac{\partial}{\partial \theta} \log f(x; \theta) \right] &= \frac{d}{d\theta} E_\theta[1] = 0 \Rightarrow \text{Cov}(W(x), \frac{\partial}{\partial \theta} \log f(x; \theta)) = \frac{d}{d\theta} E_\theta[W(x)] \\ &\Rightarrow \text{Var} \left( \frac{\partial}{\partial \theta} \log f(x; \theta) \right) = E \left[ \left( \frac{\partial}{\partial \theta} \log f(x; \theta) \right)^2 \right] - 0^2 \end{aligned}$$

$$\text{Therefore } \text{Var}(W(x)) \leq \frac{\left( \frac{d}{d\theta} E_\theta[W(x)] \right)^2}{E \left[ \left( \frac{\partial}{\partial \theta} \log f(x; \theta) \right)^2 \right]}$$

*junk*

Corollary 7.3.10 (Cramer-Rao inequality in the iid case). If the assumptions of Theorem 7.3.9 hold and, additionally,  $X_1, \dots, X_n$  are iid, then

$$\text{Var}(W(x)) \leq \frac{\left( \frac{d}{d\theta} E_\theta[W(x)] \right)^2}{n E_\theta \left[ \left( \frac{\partial}{\partial \theta} \log f(x; \theta) \right)^2 \right]}$$

*single x*

Likelihood quantities

Ex 5.6, 14/6 (b) and (c)

- likelihood  $L(\theta; x) = L(\theta) = f(x; \theta)$
- log-likelihood  $\ell(\theta) = \log L(\theta)$
- score function  $l_\theta(\theta) = u(\theta) = \frac{d\ell(\theta)}{d\theta} \rightsquigarrow \ell_{\theta_j}(\theta) = \frac{\partial \ell(\theta)}{\partial \theta_j}$
- observed information  $j(\theta) = -\frac{d^2 \ell(\theta)}{d\theta^2} = -\ell''(\theta) \rightsquigarrow j_{\theta_i \theta_j}(\theta) = -\frac{\partial^2 \ell(\theta)}{\partial \theta_i \partial \theta_j}$
- (Fisher) expected information  $i(\theta) = E\left[-\frac{d^2 \ell(\theta)}{d\theta^2}\right]$

Bertlett's identities:

- first:  $E_\theta[\ell_\theta(\theta)] = 0$
- second:  $\text{Var}_\theta[\ell_\theta(\theta)] = E_\theta[\ell_\theta(\theta)\ell_\theta(\theta)^T] = E[-\ell_{\theta\theta}(\theta)] = i(\theta)$

they hold if we can change the order of integration and derivation, see Lemma 7.3.11

Example  $x_1, \dots, x_n$   $x_i$  iid  $\sim \text{Bin}(1, p)$ 

$$f_x(x; p) = p^x (1-p)^{1-x}$$

$$L(p) = p^{\sum x_i} (1-p)^{n - \sum x_i}$$

$$\ell(p) = \sum x_i \log p + (n - \sum x_i) \log(1-p)$$

$$\frac{d\ell(p)}{dp} = \ell_p(p) = \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1-p} \quad \sum x_i - p \sum x_i = n p - p \sum x_i \quad \hat{p} = \frac{\sum x_i}{n} = \bar{x}$$

$$\ell_{pp}(p) = -\frac{\sum x_i}{p^2} - \frac{n - \sum x_i}{(1-p)^2}$$

$$i(p) = E\left[\frac{\sum x_i}{p^2} + \frac{n - \sum x_i}{(1-p)^2}\right] = \frac{n p}{p^2} + \frac{n - n p}{(1-p)^2} = \frac{n}{p} + \frac{n}{1-p} = \frac{n - n p + n p}{p(1-p)} = \frac{n}{p(1-p)}$$

$$\text{Var}(\bar{x}) = \frac{p(1-p)}{n} \rightsquigarrow \frac{1}{\frac{n}{p(1-p)}} \quad \frac{dp}{dp} = 1$$

= therefore  $\bar{x}$  is the UMVUE for  $p$

The Cramer-Rao LB is not always attainable,  $\exists$  cases in which it is smaller than any variance of unbiased estimators

Example  $X \sim \mathcal{N}(\mu, \sigma^2)$ , interested in  $\sigma^2$ , but  $\mu$  is also unknown

$$\ell_{\mu, \sigma^2}(\mu, \sigma^2) = \frac{n}{2\sigma^2} - \frac{\sum (x_i - \mu)^2}{\sigma^2}$$

$$i(\mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{n\mu^2}{\sigma^4} = \frac{n}{2\sigma^2}$$

Unbiased estimator  
 $\text{Var}(W/\mu, \sigma^2) \geq \frac{2\sigma^4}{n}$

$S^2$  estimator of  $\sigma^2$        $\text{Var}(S^2) = \frac{2\sigma^4}{n-1} \Rightarrow S^2$  does not reach the Cramer-Rao LB

Conditions to attain the Cramer-Rao LB

Corollary 7.3.15 Let  $X_1, \dots, X_n$  be iid from  $f(x; \theta)$ , which satisfies the conditions of

Theorem 7.3.9. If  $W(x)$  is any unbiased estimator of  $\tau(\theta)$ , then  $W(x)$  attains the Cramer-Rao LB iff

$$a(\theta)[W(x) - \tau(\theta)] = b(\theta) = \frac{d}{d\theta} \log L(\theta) \quad \text{for some function } a(\theta)$$

Example (cont)

$$\ell(\mu, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left( \sum (x_i - \mu)^2 \right)$$

$$\ell_{\mu, \sigma^2}(\mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{\sum (x_i - \mu)^2}{2\sigma^4} = \underbrace{-\frac{n}{2\sigma^2}}_{a(\theta)} \left( \underbrace{\frac{\sum (x_i - \mu)^2}{n}}_{\text{best unbiased estimator for } \sigma^2} - \sigma^2 \right)$$

unknown

Corollary 7.3.15 is also useful to compute the best unbiased estimator.