Lecture 8: Point estimation： Methods of Evaluating Estimators

Ma Xuejun

School of Mathematical Sciences

Soochow University

https://xuejunma.github.io

Outline

Evaluation of Estimators

- We already discussed several types of estimators and the computing issue
- We can also define any statistic to be an estimator
- Which is better? Which is worse?

Evaluation of Estimators

There are plenty of ways to evalute.Here are some popular used criteria.

- Bias and Variance
  - Unbiased estimator
  - Cramer-Rao Lower Bound
  - Rao-Blackwell Theorem
- Mean squared error (MSE)
  - Trade-off between bias and variance
  - Loss function
  - Mean squared error
- Minimax Theory
- Large sample theory
  - Consistency
  - Efficiency

Unbiasedness

■ Say that $\hat{\theta} = w(X_1, \cdots, X_n)$ is an estimator of θ,then it would be good if it satifies that

$$E[\hat{\theta}] = \theta$$

Unbiased Estimator

Let $\hat{\theta}$ be an estimator of a parameter θ.Then the bias of $\hat{\theta}$ is defined as

$$Bias(\hat{\theta}; \theta) = E_{\theta}[\hat{\theta}] - \theta$$

If $Bias(\hat{\theta}) = 0$,then we say $\hat{\theta}$ is unbias.

- $E_{\theta}[\hat{\theta}]$ means the expectation of $\hat{\theta}$ when the underlying parameter equals to θ.
- The bias is a function of θ.For unbiased estimators,the bias is a function that always equals to 0.

Unbiasedness: Example

■ Let $X_1, \cdots, X_n \sim Exp(\lambda)$. Estimate λ.
Recall that the MLE for exponential distribution is $1/\bar{X}_n$.Let the estimator be $\hat{\lambda} = 1/\bar{X}_n$.Note that $n\bar{X}_n \backsim Gamma(n, \lambda)$. Therefore,the bias is

$$Bias(\hat{\lambda}, \lambda) = \frac{n}{n-1}\lambda - \lambda = \frac{1}{n-1}\lambda$$

Therefore,the MLE $\hat{\lambda}$ is a biased estimator.However,when $n \to \infty$,the bias is close to 0.

■ Let $X_1, \cdots, X_n \sim N(\mu, \sigma^2)$.Find the bias for the sample variance.The sample variance is $\frac{1}{n-1}(X_i - \bar{X}_n)^2$.The bias is

$$Bias(\hat{\sigma}^2, \sigma^2) = E[\frac{1}{n-1}(X_i - \bar{X}_n)^2] - \sigma^2 = 0$$

So,the sample variance is unbiased estimator.

Variance

- In the previous normal example,we show that the bias for sample variance is 0.
- If we take the estimator as $\tilde{\sigma}^2 = \frac{n}{n-1}(X_1^2 - \bar{X}_n^2)$,then
$$E[\tilde{\sigma}^2] - \sigma^2 = \frac{n}{n-1}[E[X_1^2] - E[\bar{X}_n^2]] - \sigma^2 = \frac{n}{n-1} \times \frac{n-1}{n}\sigma^2 - \sigma^2 = 0,$$
which is also unbiased.
- Which estimator is better? the sample variance or $\tilde{\sigma}^2$?

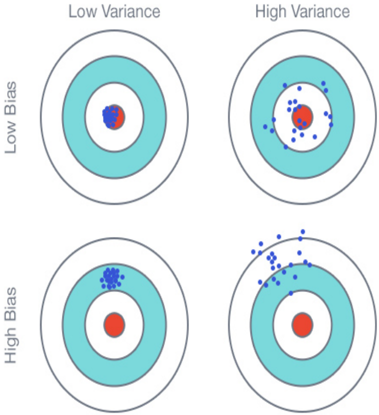
Variance

Let $\hat{\theta}$ be an estimator of the parameter θ.Then the variance of $\hat{\theta}$ is defined as

$$Var(\hat{\theta};\theta) = Var_\theta(\hat{\theta}).$$

- Targeting at θ,the estimator with smaller variance is better.
- For the previous example, the variance for sample variance is $2\sigma^4/(n-1)$,but for $\tilde{\sigma}^2$ is approximately σ^4.So,the sample variance is a better estimator.

Bias-Variance



Uniform Minimum Variance Unbiased Estimator

- Obviously,one unbiased estimator with smallest variance is the best unbiased estimator.
- However,recall that $Var(\hat{\theta};\theta)$ is a function about θ.
- It is possible that for some θ_1,$Var(\hat{\theta_1};\theta_1) < Var(\hat{\theta_2};\theta_1)$,but for another θ_2,$Var(\hat{\theta_1};\theta_2) > Var(\hat{\theta_2};\theta_2)$.
- The best unbiased estimator would be one estimator that for any other estimator W,$Var(\hat{\theta};\theta) < Var(W;\theta)$ holds for all $\theta \in \Theta$

Definition 7.3.7: Uniform Minimum Variance Unbiased Estimator

An estimator W^* of $\tau(\theta)$ is the best unbiased estimator if $E[W^*;\theta] = \tau(\theta)$ for every θ and for any other unbiased estimator W,we have

$$Var(W^*;\theta) \leq Var(W;\theta), \qquad \theta \in \Theta.$$

W^* is called the minimum variance unbiased estimator (UMVUE) for $\tau(\theta)$.

UMVUE

- Does the UMVUE exist?
    - Not necessarily.It is possible that UMVUE does not exist.
- How to prove one estimator is UMVUE?
    - There is a lower bound for the variance of unbiased estimators.If there is one unbiased estimator with variance approching the lower bound,then it is UMVUE.
- How to find the UMVUE?

Cramer-Rao Lower Bound

Theorem 7.3.9 Cramer-Rao Lower Bound

Let $X_1, \cdots, X_n$ with joint density $f(x_1, x_2, \cdots, x_n; \theta)$ and let $W : X^n \to \mathbb{R}$ be an estimator with

$$\frac{d}{d\theta}(E[W(X_1, \cdots, X_n; \theta)]) = \int \frac{\partial}{\partial\theta}[W(x_1, x_2, \cdots, x_n)f(x_1, x_2, \cdots, x_n; \theta]dx,$$

and $Var(W(X_1, \cdots, X_n); \theta) < +\infty$,then

$$Var(W(X_1, \cdots, X_n); \theta) \geq \frac{(\frac{d}{d\theta}(E[W(X_1,\cdots,X_n;\theta)]))^2}{E_\theta\{[\frac{\partial}{\partial\theta}\log(f(X_1,\cdots,X_n;\theta)))]^2\}}$$

The condition can be written as

$$\frac{d}{d\theta}(E[W(X_1, \cdots, X_n); \theta]) = \frac{d}{d\theta}\int W(x_1, x_2, \cdots, x_n)f(x_1, x_2, \cdots, x_n; \theta dx$$
$$= \int \frac{\partial}{\partial}[W(x_1, x_2, \cdots, x_n)f(x_1, x_2, \cdots, x_n; \theta]dx.$$

Remark: The integral and the derivative is exchangeable.It is satisfied under regular conditions.

Cramer-Rao Lower Bound

Corollary 7.3.10 Corollary: Unbiased Estimators

Let $X_1, \cdots, X_n$ with joint density $f(x_1, x_2, \cdots, x_n; \theta)$ and let $W : X^n \to \mathbb{R}$ be an estimator of $\tau(\theta)$.Suppose the conditions hold,then

$$Var(W;\theta) \geq \frac{\left(\frac{d}{d\theta} E[W(X_1, \cdots, X_n; \theta)]\right)^2}{nE_\theta[[\frac{\partial}{\partial\theta}\log(f(X;\theta))]^2]} = \frac{\tau'(\theta)^2}{nE_\theta[[\frac{\partial}{\partial\theta}\log(f(X;\theta))]^2]}$$

- The lower biund does not depend on the estimator.It is the lower bound for all estimators.
- The lower biund is a function of the parameter θ
- If there is an estimator W^*,which achieves the lower bound for every θ,then this estimator W^* is UMVUE.
- No need to prove $Var(W^*;\theta) \leq Var(W;\theta)$ for all W.

Score and Fisher Information

- An important item here is $E_\theta[[\frac{\partial}{\partial\theta}\log(f(X_1,\cdots,X_n;\theta)))]^2]$
- Actually,we hve some notions and lemmas w.r.t.this quantity

Score function

Let $X_1,\cdots,X_n$ be with joint density $f(x_1,x_2,\cdots,x_n;\theta)$ .The score function is the derivative of the log-likelihood function,which is

$$S_n(\theta)=\frac{\partial}{\partial\theta}\log(f(X_1,\cdots,X_n;\theta))$$

If $X_1,\cdots,X_n$ are i.i.d.with density $f(x;\theta)$,then the score function equals to

$$\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n}f(X_i;\theta)=\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(X_i;\theta)$$

Score function

Lemma

Under regularity conditions,

$$E_\theta[S_n(\theta)] = 0$$

Proof The expectation of score function is

$$\begin{aligned}
E_\theta[S_n(\theta)] &= \int \frac{\partial \log(f(x_1,\cdots,x_n;\theta)}{\partial\theta} f(x_1,\cdots,x_n;\theta)dx_1\cdots dx_n \\
&= \int \frac{\frac{\partial}{\partial\theta} f(x_1,\cdots,x_n;\theta)}{f(x_1,\cdots,x_n;\theta)} f(x_1,\cdots,x_n;\theta)dx_1\cdots dx_n \\
&= \int \frac{\partial}{\partial\theta} f(x_1,\cdots,x_n;\theta)dx_1\cdots dx_n \\
&= \frac{\partial}{\partial\theta}\int f(x_1,\cdots,x_n;\theta)dx_1\cdots dx_n = \frac{\partial}{\partial\theta}1 = 0
\end{aligned}$$

Note If θ mismatches,it may not hold.It is possible $E_{\theta_1}[S_n(\theta_2)] \neq 0$

Fisher Information

Fisher Information

Let $X_1, \cdots, X_n$ be with with joint density $f(x_1, x_2, \cdots, x_n; \theta)$. The Fisher information is the variance of the score function,which is

$$I_n(\theta) = \mathbf{Var}_\theta(S_n(\theta)) = E[[\frac{\partial}{\partial\theta} \log f(X_1, \cdots, X_n; \theta)]^2]$$

If $X_1, \cdots, X_n$ are i.i.d.with density $f(x; \theta)$,then the Fisher information is

$$I_n(\theta) = nI(\theta)$$

where $I(\theta)$ is the Fisher information for single observation.

- $S_n(\theta) = 0$,yet $Var_\theta(S_n(\theta))$ is a function of θ
- **Proof** in i.i.d.case,the score function is $S_n(\theta) = \sum_{i=1}^{n} \log f(X_i; \theta)$

$$I_n(\theta) = Var_\theta(S_n(\theta)) = Var(\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i; \theta))$$

$$= \sum_{i=1}^{n} Var(\frac{\partial}{\partial\theta} \log f(X_i; \theta)) = nVar(\frac{\partial}{\partial\theta} \log f(X_1; \theta)) = nI(\theta)$$

Fisher Information

- According to the Cramer-Rao lower bound,all the unbiased estimator for θ has variance larger than $1/I_n(\theta)$.So $I_n(\theta)$ gives us the bound for the information we can get from the data. That's why we call is as Information.
- Another statement of Cramer-Rao lower bound

Corollary:Unbiased Estimators

Let $X_1, \cdots, X_n$ be i.i.d.samples with density $f(x;\theta)$ and let $W : X^n \to \mathbb{R}$ be an unbiased estimator of $\tau(\theta)$.Suppose the conditions hold,then

$$Var(W;\theta) \geq \frac{\tau'(\theta)^2}{I_n(\theta)} = \frac{\tau'(\theta)^2}{nI(\theta)}$$

- Obviously,the variance will converge to 0 when n increases.
- The best unbiased estimator has convergence rate at $1/\sqrt{n}$.

Fisher Information

Lemma: Fisher Information

Under regularity conditions,

$$I_n(\theta) = E\Big\{[\tfrac{\partial}{\partial\theta}\log f(X_1,\cdots,X_n;\theta)]^2\Big\} = -E[\tfrac{\partial^2}{\partial\theta^2}\log f(X_1,\cdots,X_n;\theta)]$$

Proof. In short,we denote $X=(X_1,X_2,\cdots,X_n)$.For the L.H.S,there is

$$E[[\tfrac{\partial}{\partial\theta}\log f(X;\theta)]^2] = E[\tfrac{1}{(f(X;\theta))^2}(\tfrac{\partial}{\partial\theta}f(X;\theta))^2]$$

For the R.H.S,we have

$$\begin{aligned}
-E\Big[\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\Big] &= -E\Big[\frac{\partial}{\partial\theta}\frac{1}{f(X;\theta)}\frac{\partial f(X;\theta)}{\partial\theta}\Big] \\
&= E\Big[\frac{1}{(f(X;\theta))^2}[\frac{\partial}{\partial\theta}f(X;\theta)]^2\Big] - E\Big[\frac{1}{f(X;\theta)}\frac{\partial f(X;\theta)}{\partial\theta}\Big] \\
&= E\Big[\frac{1}{(f(X;\theta))^2}[\frac{\partial}{\partial\theta}f(X;\theta)]^2\Big] - \int\frac{\partial^2 f(X;\theta)}{\partial\theta^2}dx \\
&= L.H.S - \frac{\partial^2}{\partial\theta^2}\int f(X;\theta)dx^1 = L.H.S
\end{aligned}$$

[1] true for an exponential family

- **Theorem 7.3.9 (Cramér-Rao Inequality)** Let $X_1, X_2, \cdots, X_n$ be a sample with pdf $f(\boldsymbol{x}|\theta)$, and let $W(\mathbf{X})$ be any estimator satisfying

$$\frac{d}{d\theta}E_\theta W(\boldsymbol{X}) = \int_{\mathcal{X}} \frac{\partial}{\partial\theta}\left[W(\boldsymbol{x})f(\boldsymbol{x}|\theta)\right] d\boldsymbol{x}$$

and $\text{Var}_\theta W(\boldsymbol{X}) < \infty$. Then

$$\text{Var}_\theta W(\boldsymbol{X}) \geq \frac{\left(\frac{d}{d\theta}E_\theta W(\boldsymbol{X})\right)^2}{E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}|\theta)\right)^2\right)}$$

In particular, if $W(X)$ is an unbiased estimator of θ, then

$$\text{Var}_\theta W(\boldsymbol{X}) \geq \frac{1}{E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}|\theta)\right)^2\right)}$$

- **Corollary 7.3.10 (Cramér-Rao Inequality, iid case)** If X_1, X_2, $\cdots, X_n$ are i.i.d. $f(x|\theta)$, and the condition of Theorem 7.3.9 are satisfied, then

$$\text{Var}_\theta\left(W(\boldsymbol{X})\right) \geq \frac{\left(\frac{d}{d\theta}E_\theta W(\boldsymbol{X})\right)^2}{nE_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right)}$$

Example of CRLB-Poisson

Suppose $X_1, \cdots, X_n$ from a iid sample from Poisson distribution,

$$f(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Find the CRLB for $\hat{\lambda}$.
Solution For the Poisson distribution,

$$l(\lambda) = X \ln \lambda - \lambda - \ln X!$$
$$l'(\lambda) = \frac{X}{\lambda} - 1 \qquad l''(\lambda) = \frac{X}{\lambda^2}$$
$$I(\lambda) = \frac{E[X]}{\lambda^2} = \frac{1}{\lambda}$$

Finally,we have the CRLB $\frac{\lambda}{n}$.
Recall that the MLE for Poisson example is $\bar{X}_n$,with expectation λ and variance $\frac{\lambda}{n}$.So the MLE is UMVUE for Poisson distribution.

- **Example 7.3.12** $\bar{X}$ is UMVUE for λ if $X_1, \cdots, X_n$ are i.i.d. Poisson(λ). From Theorem 7.3.9, we have for any unbiased estimator $W(\mathbf{X})$ of λ.

$$
\begin{aligned}
\text{Var}_\lambda W(\boldsymbol{X}) &\geq \frac{1}{-nE_\lambda\left[\frac{\partial^2}{\partial\lambda^2}\log f(\boldsymbol{x}|\lambda)\right]} &\quad (4.1)\\
\log f(\boldsymbol{x}|\lambda) &= \log\left[e^{-\lambda}\frac{\lambda^x}{x!}\right] = -\lambda + x\log\lambda - \log x! \\
\frac{\partial^2}{\partial\lambda^2}\log f(\boldsymbol{x}|\lambda) &= -x\frac{1}{\lambda^2}.
\end{aligned}
$$

Therefore, $-E_\lambda\left[\frac{\partial^2}{\partial\lambda^2}\log f(\boldsymbol{x}|\lambda)\right] = \frac{1}{\lambda^2}E_\lambda X = \frac{1}{\lambda}$.
(10.1) Becomes $\text{Var}_\lambda(W(\boldsymbol{X})) \geq \frac{\lambda}{n}$.
But $\text{Var}_\lambda(\bar{X}) = \frac{\lambda}{n}$.

Example of CRLB-Normal

Example

Let $X_1, \cdots, X_n$ be a random sample from the $N(\mu, \sigma^2)$ distribution.Find the CRLB and,in case 1 and 2.check whether it is equalled,for the variance of an biased estimator of

- μ when σ^2 is known,
- σ^2 whenμ is known
- μ when σ^2 is unknown
- σ^2 whenμ is unknown

Example of CRLB-Normal

Solution:The sample joint pdf is

$$f_X(X|\theta) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp(-\tfrac{1}{2}(x_i-\mu)^2/\sigma^2)$$

and

$$\log f_X(X|\theta) = -\tfrac{n}{2}\log(2\pi) - \tfrac{n}{2}\log(\sigma^2) - \tfrac{1}{2} = \sum_{i=1}^{n}(x_i-\mu)^2/\sigma^2$$

1.When σ^2 is known $\theta = \mu$ and

$$\log f_X(X|\theta) = -\tfrac{n}{2}\log(2\pi) - \tfrac{n}{2}\log(\sigma^2) - \tfrac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2/\sigma^2$$

$$S(X) = \tfrac{\partial}{\partial\theta}\log f_X(X|\theta) = \sum_{i=1}^{n}(x_i-\mu)^2/\sigma^2 = \tfrac{n}{\sigma^2}[\bar{x}-\theta]$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$,is a n unbiased estimator of $\theta = \mu$ whose variance equals the CRLB and that $\frac{n}{\sigma^2} = I(\theta)$ i.e.CRLB=$\frac{\sigma^2}{n}$.Thus $\bar{X}$ is UMVUE.

Example of CRLB-Normal

2.When μ is known but σ^2 is unknown,$\theta = \sigma^2$ and

$$\log f_X(X|\theta) = -\tfrac{n}{2}\log(2\pi) - \tfrac{n}{2}\log(\theta) - \tfrac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu)^2/\sigma^2$$

Hence

$$S(X) = \frac{\partial}{\partial\theta}\log f_X(X|\theta) = -\frac{n}{2\theta} + \frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu)^2/\sigma^2$$

$$= \frac{n}{2\theta^2}[\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2 - \theta]$$

$\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2$ is an unbiased estimator of $\tau = \sigma^2$ and $\frac{n}{2\theta^2} = I(\theta)$ i.e.the $CRLB = \frac{2\theta^2}{n} = \frac{2\sigma^4}{n}$

Example of CRLB-Normal

3.and 4. Case both μ and σ^2 is unknown
here $\theta = \binom{\mu}{\sigma^2} = \binom{\theta_1}{\theta_2}$ i.e.$\theta_1 = \mu$ and $\theta_2 = \sigma^2$

$$f_X(X|\theta) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{1}{2}(x_i - \mu)^2/\sigma^2) \quad \propto \theta_2^{-n/2} \exp(\frac{1}{2\theta_2} \sum_{i=1}^{n} (x_i - \theta_1)^2/\sigma$$

and

$$\log f_X(X|\theta) = -\tfrac{n}{2} \log \theta_2 - \tfrac{1}{2\theta_2} \sum_{i=1}^{n} (x_i - \theta_1)^2/\sigma^2$$

Thus

$$= \tfrac{\partial}{\partial \theta} \log f_X(X|\theta) = \tfrac{1}{\theta_2} \sum_{i=1}^{n} (x_i - \theta_1)/\sigma^2$$

$$\tfrac{\partial^2}{\partial \theta^2} \log f_X(X|\theta) = -\tfrac{n}{\theta_2}$$

$$\tfrac{\partial^2}{\partial \theta^2 \theta^1} \log f_X(X|\theta) = -\tfrac{1}{\theta_2^2} \sum_{i=1}^{n} (x_i - \theta_1)$$

$$\tfrac{\partial}{\partial \theta^2} \log f_X(X|\theta) = -\tfrac{n}{2\theta_2} + \tfrac{1}{2\theta_2^2} \sum_{i=1}^{n} (x_i - \theta_1)^2$$

Example of CRLB-Normal

$$\frac{\partial^2}{\partial\theta^2}\log f_X(X|\theta) = \frac{n}{2\theta_2^2} - \frac{1}{\theta_2^3}\sum_{i=1}^{n}(x_i-\theta_1)^2$$

Consequently

$$I_{11}(\theta) = -E(-\frac{n}{\theta_2}) = \frac{n}{\theta_2}$$

$$I_{12}(\theta) = -E(-\frac{1}{\theta_2^2}\sum_{i=1}^{n}(x_i-\theta_1)) = 0$$

$$I_{22}(\theta) = -E(\frac{n}{2\theta_2} - \frac{1}{\theta_2^3}\sum_{i=1}^{n}(x_i-\theta_1)^2) = \frac{n}{2\theta_2^2}$$

Example of CRLB-Normal

i.e

$$I(\theta) = \begin{bmatrix} \frac{n}{\theta_2} & 0 \\ 0 & \frac{n}{2\theta_2^2} \end{bmatrix}$$

and

$$[I(\theta)]^{-1} = J(\theta) = \begin{bmatrix} \frac{\theta_2}{n} & 0 \\ 0 & \frac{2\theta_2^2}{n} \end{bmatrix} = \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix}$$

Consequently,for unbiased estimators $\hat{\mu}, \hat{\sigma^2}$ of μ and σ^2 respectively

$$Var(\hat{\mu}) \geq \frac{\sigma^2}{n}$$

and

$$Var(\hat{\sigma^2}) \geq \frac{2\sigma^4}{n}$$

The vector case

■ For the normal example,e=we consider $\theta = (\mu, \sigma^2)$,where the unknown parameter is a vector.

■ Let $\theta = (\theta_1, \cdots, \theta_k)$,then the score function is

$$S_n(\theta) = (\frac{\partial}{\partial\theta_1 l(\theta)}, \frac{\partial}{\partial\theta_2 l(\theta)}, \cdots, \frac{\partial}{\partial\theta_k l(\theta)})^T$$

$E[S_n(\theta)] = 0$ still holds.

■ The Fisher information is now a $k \times k$ matrix,actually,the covariance matrix for $S_n(\theta)$,that

$$I_n = E_\theta[S_n(\theta)(S_n(\theta))^T],$$

For the (r, s) element of I_n,there is $I_n(r, s) = -E_\theta[\frac{\partial^2 l(\theta)}{\partial\theta_r \partial\theta_s}]$.So,under regular conditions,I_n equals to the expectation of the Hessian matrix for $-l(\theta)$.

Methods of Evaluating Estimators

The mean square error (MSE) of an estimator W of a parameter θ is the function of θ defined by

$$E_\theta(W-\theta)^2.$$

$\text{Bias}_\theta W = E_\theta W - \theta$. If $\text{Bias}_\theta W = 0$, then W is unbiased.

- Example 7.3.3 (Normal MSE) Let $X_1, X_2, \cdots, X_n$ be i.i.d. $N(\mu, \sigma^2)$. Then statistics $\bar{X}$ and S^2 are both unbiased.

$$\begin{aligned} MSE(\bar{X}) &= E(\bar{X}-\mu)^2 = \text{Var}(\bar{X}) = \sigma^2/n \\ E(S^2-\sigma^2)^2 &= \text{Var}(S^2) = \frac{2\sigma^4}{n-1}{}^{2} \end{aligned}$$

---

[2]

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \Rightarrow \text{Var}\Big(\frac{(n-1)S^2}{\sigma^2}\Big) = 2(n-1)$$

- Example 7.3.4 Maximum Likelihood estimator of σ^2 is $\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{n-1}{n}S^2$.

$$
\begin{aligned}
\mathrm{Var}\left(\frac{n-1}{n}S^2\right) &= \frac{(n-1)^2}{n^2}\cdot\frac{2\sigma^4}{n-1} = \frac{2(n-1)}{n^2}\sigma^4 \\
MSE\left(\frac{n-1}{n}S^2\right) &= \left(\frac{n-1}{n}ES^2-\sigma^2\right)^2 + \frac{2(n-1)}{n^2}\sigma^4 \\
&= \sigma^4\left(\frac{n-1}{n}-1\right)^2 + \frac{2(n-1)}{n^2}\sigma^4 \\
&= \sigma^4\frac{2n-1}{n^2}
\end{aligned}
$$

Since

$$\frac{2n-1}{n^2} < \frac{2}{n-1},$$

So in this case MLE has smaller MSE than the unbiased estimator S^2.

Remark While MSE is a reasonable measurement for location parameters, it may not be a good to compare estimators of scale parameters with MSE.

Example 7.3.5 (MSE of binomial Bayes Estimator) $X_1, \cdots, X_n \sim$ Bernoulli(p).

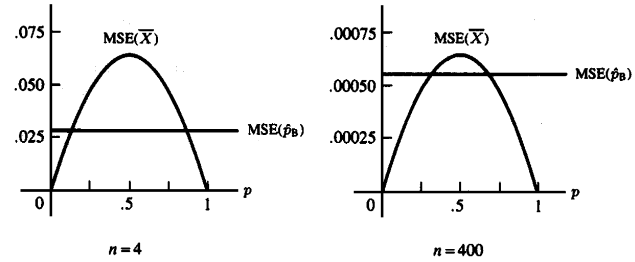
- Let $\hat{p} = \frac{X_1 + \cdots + X_n}{n}$. $E_p(\hat{p} - p)^2 = \text{Var}_p(\bar{X}) = \frac{p(1-p)}{n}$.
- Let $\hat{p}_B = \frac{Y+\alpha}{\alpha+\beta+n}$ be the Bayes estimator. Here $Y = \sum_{i=1}^n X_i$

$$\begin{aligned} MSE(\hat{p}) &= \text{Var}_p(\hat{p}_B) + (\text{Bias}_p(\hat{p}_B))^2 \\ &= \text{Var}\left(\frac{Y+\alpha}{\alpha+\beta+n}\right) + \left(E_p\left(\frac{Y+\alpha}{\alpha+\beta+n}\right) - p\right)^2 \\ &= \frac{np(1-p)}{(\alpha+\beta+n)^2} + \left(\frac{np+\alpha}{\alpha+\beta+n} - p\right)^2 \end{aligned}$$

In the absence of good prior information about p, we might choose α and β to make the MSE of $\hat{p}_B$ constant. Choose $\alpha = \beta = \sqrt{n/4}$ gives

$$\hat{p}_B = \frac{Y + \sqrt{n/4}}{n + \sqrt{n}}, \; E(\hat{p}_B - p)^2 = \frac{n}{4(n+\sqrt{n})^2}$$

Figure 7.3.1 Comparison of $MSE(\hat{p})$ and $MSE(\hat{p}_B)$ for sample size $n = 4$ and $n = 400$ in Example 7.3.5



- For small n, $\hat{p}_B$ is the better choice (unless there is a strong belief that p is near 0 or 1)
- For large n, $\hat{p}$ is the better choice (unless there is a strong belief that p is close to $\frac{1}{2}$)

Best Unbiased Estimators

As we have discussed, there is usually no "best MSE" estimator. However, if we restrict our choice from unbiased estimators, then there exists best estimator in this class.

Definition 7.3.7 An estimator W^* is a best unbiased estimator of $\tau(\theta)$ if it satisfies $E_\theta W^* = \tau(\theta)$ for all θ and, for any other estimator W with $E_\theta W = \tau(\theta)$, we have

$$\text{Var}_\theta W^* \leq \text{Var}_\theta W \quad \text{for all } \theta.$$

W^* is also called a uniform minimum variance unbiased estimator (UMVUE) of $\tau(\theta)$.

* Finding UMVUE is not easy.

To evaluate $E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right)$, we have the following Lemma.

- **Lemma 7.3.11** If $f(x|\theta)$ satisfies

$$\frac{d}{d\theta}E_\theta\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right) = \int \frac{\partial}{\partial\theta}\left[\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right) f(x|\theta)\right] dx$$

(true for an exponential family), then

$$E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right) = -E_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right).$$

- **Example 7.3.12** $\bar{X}$ is UMVUE for λ if $X_1, \cdots, X_n$ are i.i.d. Poisson(λ). From Theorem 7.3.9, we have for any unbiased estimator $W(\mathbf{X})$ of λ.

$$\begin{aligned} \text{Var}_\lambda W(\boldsymbol{X}) &\geq \frac{1}{-nE_\lambda\left[\frac{\partial^2}{\partial\lambda^2}\log f(\boldsymbol{x}|\lambda)\right]} \qquad (10.1) \\ \log f(\boldsymbol{x}|\lambda) &= \log\left[e^{-\lambda}\frac{\lambda^x}{x!}\right] = -\lambda + x\log\lambda - \log x! \\ \frac{\partial^2}{\partial\lambda^2}\log f(\boldsymbol{x}|\lambda) &= -x\frac{1}{\lambda^2}. \end{aligned}$$

Therefore, $-E_\lambda\left[\frac{\partial^2}{\partial\lambda^2}\log f(\boldsymbol{x}|\lambda)\right] = \frac{1}{\lambda^2}E_\lambda X = \frac{1}{\lambda}$.
(10.1) Becomes $\text{Var}_\lambda(W(\boldsymbol{X})) \geq \frac{\lambda}{n}$.
But $\text{Var}_\lambda(\bar{X}) = \frac{\lambda}{n}$.

- **Example 7.3.13 (Unbiased Estimator for Scale Parameter)** Let $X_1, \cdots, X_n$ be i.i.d. with pdf $f(x|\theta) = \frac{1}{\theta}$, $0 < x < \theta$. Since $\frac{\partial}{\partial \lambda} \log f(x|\theta) = -\frac{1}{\theta}$, we have

$$E_\theta \left[\frac{\partial}{\partial \lambda} \log f(x|\theta) \right] = \frac{1}{\theta^2}$$

  So if W is unbiased for θ, then

$$\text{Var}_\theta(W) \geq \frac{\sigma^2}{n}.$$

- On the other hand, $Y = \max(Y_1, \cdots, Y_n)$ is a sufficient statistic. $f_Y(y|\theta) = ny^{n-1}/\theta^n$, $0 < y < \theta$. So

$$E_\theta Y = \int_0^\theta y \cdot \frac{ny^{n-1}}{\theta^n} dy = \frac{n}{n+1}\theta,$$

  showing that $\frac{n+1}{n} Y$ is an unbiased estimator of θ.

$$
\begin{aligned}
\mathrm{Var}_\theta\left(\frac{n+1}{n}Y\right) &= \left(\frac{n+1}{n}\right)^2 \mathrm{Var}_\theta(Y) \\
&= \left(\frac{n+1}{n}\right)^2 \left[E_\theta Y^2 - (EY)^2\right] \\
&= \left(\frac{n+1}{n}\right) \left[\frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2\right] \\
&= \frac{1}{n(n+2)}\theta^2,
\end{aligned}
$$

which is uniformly smaller than θ^2/n. Cramér-Rao lower bound Theorem is not applicable to this pdf since $\frac{d}{d\theta}\int_0^\theta h(x)f(x|\theta)dx \neq \int_0^\theta h(x)\frac{\partial}{\partial\theta}f(x|\theta)dx.$

- **Example 7.3.14 (Normal Variance Bound)** Let $X_1, \cdots, X_n$ be i.i.d. $N(\mu, \sigma^2)$. The conditions of Cramér-Rao Theorem are satisfied. Let W be an unbiased estimator of σ^2, then

$$\text{Var}(W|\mu, \sigma^2) \geq 2\sigma^4/n.$$

In Example 7.3.3 we see that $\text{Var}(S^2|\mu, \sigma^2) \geq \frac{2\sigma^4}{n-1}$. So S^2 does not attain the Cramér-Rao lower bound.

- **Corollary 7.3.15 (Attainment)** Let $X_1, \cdots, X_n$ be i.i.d. $f(x|\theta)$, where $f(x|\theta)$ satisfies the conditions of the Cramér-Rao Theorem. Let $L(\theta|\boldsymbol{x})$ denote the likelihood function. If $W(\mathbf{X})$ is any unbiased estimator of $\tau(\theta)$, then $W(\mathbf{X})$ attains the Cramér-Rao lower bound if and only if

$$a(\theta)[W(\boldsymbol{x}) - \tau(\theta)] = \frac{\partial}{\partial\theta} \log L(\boldsymbol{x}|\theta)$$

for some function $a(\theta)$.

- **Proof** The Cramér-Rao inequality, can be written as

$$\left[\mathrm{Cov}_\theta\left(W(\mathbf{X}), \frac{\partial}{\partial\theta}\log\prod_{i=1}^n f(X_i|\theta)\right)\right]^2$$

$$\leq \quad \mathrm{Var}_\theta W(\mathbf{X}) \cdot \mathrm{Var}_\theta\left(\frac{\partial}{\partial\theta}\log L(\mathbf{X})\right)$$

Using the condition for "=" in Cauchy-Schwarz inequality, we obtain the expression (10.1).

- **Example 7.3.16 (Continuation of Example 7.3.14)**

$$L(\boldsymbol{x}|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2/\sigma\right)$$

and hence

$$\frac{\partial}{\partial\sigma^2}\log L(\boldsymbol{x}|\mu,\sigma^2) = \frac{n}{2\sigma^4}\left(\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{n} - \sigma^2\right)$$

Taking $a(\sigma^2) = \frac{n}{2\sigma^4}$ shows that the best unbiased estimator of σ^2 is $\sum_{i=1}^{n}(x_i-\mu)^2/n$, which is calculable only if μ is known.

- So the question of finding best unbiased estimator are still unsolved for many common pdf's.