Ch 7. Point estimation

Intro

Definition

A function of random variables $X_1, \cdots, X_n$ is called *Statistic*.

Definition

The probability distribution of a statistic T is called the sampling distribution of T.

$\triangleright$ Example: $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Let $T = \bar{X}$, then sampling distribution of T is

$$N(\mu, \sigma^2/n).$$

If $T = S^2$ then sampling distribution of S^2 is

$$(n-1)S^2/\sigma^2 \sim \chi^2_{(n-1)}.$$

Definition

An *estimator* is a function of random variables $X_1, \cdots, X_n$, $T = W(X_1, \cdots, X_n)$.

$\triangleright$ Note:

1. Estimator is actually a statistic.
2. Estimator is also random.
3. An *estimate* is a function of realized values of $X_1 = x_1, \cdots, X_n = x_n$. $t = W(x_1, \cdots, x_n)$

$\triangleright$ Example: $T = \bar{X}$, $\hat{F}_n(x_0) = n^{-1}\sum_{i=1}^{n} I[X_i \leq x_0]$, $T = (\bar{X}, S^2)$.

How to estimate the parametric function $\tau(\boldsymbol{\theta})$ using the random sample $X_1, \cdots, X_n$? : MME, MLE, BE so on

Let $X_1, \cdots, X_n \overset{iid}{\sim} f(x|\boldsymbol{\theta})$, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k) \in \Theta$.

Definition

$$j^{th} \text{ Population moment: } \mu_j(\boldsymbol{\theta}) = E\left(X^j\right)$$

$$j^{th} \text{ Sample moment: } m_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j$$

▷ Note:

1. μ_j is a function of $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)$.
2. $E[m_j] = E\left[n^{-1}\sum_{i=1}^n X_i^j\right] = \mu_j$.

Definition

MME of $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)$, denoted by $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1, \cdots, \tilde{\theta}_k)$, is defined as a solution of the system of equations

$$m_j = \mu_j(\theta_1, \cdots, \theta_k), \; j = 1, \cdots .l$$

▷ Example: $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Find MME of $\boldsymbol{\theta} = (\mu, \sigma^2)$.

▷ Example: $X_1, \cdots, X_n \overset{iid}{\sim}$ Gamma(α, β). Find MME of $\boldsymbol{\theta} = (\alpha, \beta)$.

▷ Example: $X_1, \cdots, X_n \overset{iid}{\sim}$ Beta(α, β). Find MME of $\boldsymbol{\theta} = (\alpha, \beta)$.

- Note:
  1. MM equations may have multiple solutions or no solution. The solution may fall outside of the parameter space.
  2. MME may not be applicable if the population moments do not exist such as Cauchy distribution.
  3. One may not successful considering the first k–moments.
     - Example: $X_1, \cdots, X_n \overset{iid}{\sim} f(x|\beta)$, where

$$f(x|\beta) = \frac{1}{2\beta} e^{-|x|/\beta}.$$

Definition

Let $f(x_1, \cdots, x_n|\boldsymbol{\theta})$ be the joint pdf/pmf of $X_1, \cdots, X_n$. For a fixed $x_1, \cdots, x_n$,

$$L(\boldsymbol{\theta}) = f(x_1, \cdots, x_n|\boldsymbol{\theta})$$

as a function of $\boldsymbol{\theta}$, is called the likelihood function. $\ln[L(\boldsymbol{\theta})]$ is called the log likelihood function.

With discrete random variable,

$$L(\boldsymbol{\theta}) = P(X_1 = x_1, \cdots, X_n = x_n).$$

$X_1, \cdots, X_n \overset{iid}{\sim} f(x|\boldsymbol{\theta})$. $L(\boldsymbol{\theta}) =$

Definition

Let $f(x_1, \cdots, x_n|\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$ be the joint pdf/pmf of $X_1, \cdots, X_n$. Then for a given set of observations $(x_1, \cdots, x_n)$, the *maximum likelihood estimate* of $\boldsymbol{\theta}$ is a point $\boldsymbol{\theta}_0 = h(x_1, \cdots, x_n)$ satisfying

$$f(x_1, \cdots, x_n|\boldsymbol{\theta}_0) = \max_{\theta \in \Theta} f(x_1, \cdots, x_n|\boldsymbol{\theta}).$$

The *maximum likelihood estimator (MLE)* is defined as

$$\hat{\boldsymbol{\theta}} = h(X_1, \cdots, X_n).$$

How to find MLE ?

- Using differentiation, - Direct maximization, - Numerical evaluation

- Assume $L(\theta)$ is twice differentiable in the interior points of Θ. Then $\hat{\theta}$ maximizes $L(\theta)$ if
  1. $\hat{\theta}$ is the unique value satisfying

$$\frac{dL(\theta)}{d\theta} \left(\frac{d \ln[L(\theta)]}{d\theta} \right) \Bigg|_{\hat{\theta}} = 0$$

$$\frac{d^2 L(\theta)}{d\theta^2} \left(\frac{d^2 \ln[L(\theta)]}{d\theta^2} \right) \Bigg|_{\hat{\theta}} < 0$$

  2. The maximizer does not occur at the boundary of the parameter space.

- If Θ is open, then the unique value of $\hat{\theta}$ satisfying (1) is the MLE.
- For multidimensional parameter space $\Theta \in R^k$ when $L[\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)]$ has partial derivatives with respect to θ_i's, then differentiate $L(\boldsymbol{\theta})$ or equivalently $\ln[L(\boldsymbol{\theta})]$ to find the MLE. That is, the solutions for

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_i} \left(\frac{\partial \ln[L(\boldsymbol{\theta})]}{\partial \theta_i} \right) = 0, \quad i = 1, \cdots, k$$

are the mle of $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)$

$\triangleright$ Example: $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Known σ and Unknown σ
[See Examples 7.2.5 and 7.2.11]

$\triangleright$ Example: $X_1, \cdots, X_n \overset{iid}{\sim} Bernoulli(p)$.
[See Example 7.2.7. In addition, see Example 7.2.9 for binomial(n, p) with known p, unknown n.]

Theorem

Let $\hat{\boldsymbol{\theta}}$ be the mle of $\boldsymbol{\theta}$. Then for any function $\tau(\boldsymbol{\theta})$, the mle of $\tau(\boldsymbol{\theta})$ is defined to be $\tau(\hat{\boldsymbol{\theta}})$.

▷ Example: $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Find mle of e^μ, μ^2, σ/μ and $P[X \leq a]$.

▷ Note:

1. It is possible that the likelihood equations do not have closed-form solution. May need a numerical method.
2. When the likelihood function is not differentiable, we may maximize $L(\boldsymbol{\theta})$ directly.

▷ Example: $X_1, \cdots, X_n \overset{iid}{\sim} \text{Gamma}(\alpha, \beta)$. (No closed form. Iterative approach is required.)

▷ Example: $X_1, \cdots, X_n \overset{iid}{\sim} f(x|\theta)$, where

$$f(x|\theta) = e^{-(x-\theta)}, \ \theta \leq x < \infty$$

▷ Example: $X_1, \cdots, X_n \overset{iid}{\sim} f(x|\theta)$, $f(x|\theta) = \frac{1}{2}e^{-|x-\theta|}, \ x \in R$

$$ln[L(\theta)] = -n \log 2 - \sum_{i=1}^{n} |x_i - \theta|.$$

The first partial derivative is $\sum_{i=1}^{n} sgn(x_i - \theta)$, where $sgn(t)$=1,0, or -1 depending on whether $t > 0$, t=0, or $t < 0$. Note that we have used $\frac{d}{dt}|t| = sgn(t)$, which is true unless $t = 0$. Thus, $\hat{\theta}^{MLE}$ is median of $\{x_1, x_2, \ldots, x_n\}$.

$\triangleright$ Example: $X_{ij}, \ i = 1, \cdots, s; \ j = 1, \cdots, n$ independently distributed as normal distribution with mean μ_i and variance σ^2. Find the mle of μ_i and σ^2.

So far, $X_1, \cdots, X_n \overset{iid}{\sim} f(x|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is **unknown** and is assumed **fixed**. In **Bayesian** framework, $\boldsymbol{\theta}$ is assumed **random**. The distribution of $\boldsymbol{\theta}$ is called **prior distribution**, denoted by

$$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

▷ Example: Consider a machine which makes parts for cars. θ: fraction of defective. On a certain day, 10 pieces are examined.

$$X_i = \begin{cases} 1, & \text{if } i\text{th piece is defective,} \\ 0, & \text{otherwise,} \end{cases}$$

$i = 1, \cdots, 10$. MME or MLE? (ans: $\bar{X}$.)

$\triangleright$ Example -continued: Now assume that mechanic knows something about θ and gives a statistical model for θ

$$\pi(\theta) = 6\theta(1-\theta),\ 0 \leq \theta \leq 1.$$

Prior distribution of θ is Beta(2,2) distribution.

- In Bayesian frame, one of the goal of the inference about θ is to find a posterior distribution of θ.
- Then, how should we use the data $X_1 = x_1, \cdots, X_n = x_n$ to achieve the goal? (Use Bayes' Theorem.)

The conditional distribution of $\boldsymbol{\theta}$ conditioning on $X_1 = x_1, \cdots, X_n = x_n$ is called **posterior distribution** of $\boldsymbol{\theta}$.

$$
\begin{aligned}
\pi(\boldsymbol{\theta}|x_1, \cdots, x_n) &= \frac{f(x_1, \cdots, x_n, \boldsymbol{\theta})}{m(x_1, \cdots, x_n)} \\
&= \frac{\pi(\boldsymbol{\theta}) f(x_1, \cdots, x_n|\boldsymbol{\theta})}{\sum_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \pi(\boldsymbol{\theta}) f(x_1, \cdots, x_n|\boldsymbol{\theta})} \\
&= \frac{\pi(\boldsymbol{\theta}) f(x_1, \cdots, x_n|\boldsymbol{\theta})}{\int_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \pi(\boldsymbol{\theta}) f(x_1, \cdots, x_n|\boldsymbol{\theta}) d\boldsymbol{\theta}} \\
&\propto \pi(\boldsymbol{\theta}) f(x_1, \cdots, x_n|\boldsymbol{\theta})
\end{aligned}
$$

- Any Bayesian inference is based on this posterior distribution of $\boldsymbol{\theta}$.

▷ Example-Continued: $X_1, \cdots, X_n \overset{iid}{\sim}$ Bernoulli(θ), $\theta \sim$ Beta(α, β).

$$f(x_1, \cdots, x_n|\theta) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

α and β are constants and known.

$\pi(\theta|x_1, \cdots, x_n) \propto$

- Bayes estimator with respect to squared error loss is:
- We see the Bayes estimator is a weighted average.

Definition

Let $f(x_1, \cdots, x_n|\boldsymbol{\theta})$ denote the pdf/pmf for the data $X_1, \cdots, X_n$. The prior distribution $\pi(\boldsymbol{\theta})$ belongs to family $\Pi = \{\pi(\boldsymbol{\theta})\}$. If the posterior $\pi(\boldsymbol{\theta}|x_1, \cdots, x_n)$ is also in Π then Π (or the distribution of $\boldsymbol{\theta}$) is said to be a *conjugate family* [for the distribution $f(x_1, \cdots, x_n|\boldsymbol{\theta})$.]

$\triangleright$ Example: $X|\theta \sim N(\theta, \sigma^2)$, σ^2 is known. $\theta \sim N(\mu, \tau^2)$, μ and τ^2 are known. Find the posterior distribution of θ.

$\triangleright$ Example: Find the posterior distribution for the following prior and likelihood.

- $X_1, \cdots, X_n \overset{iid}{\sim}$ Poisson(θ), $\theta \sim$ Gamma(α, β)
- $X_1, \cdots, X_n \overset{iid}{\sim}$ Bernoulli(p), $p \sim$ Beta(α, β)
- $X_1, \cdots, X_n \overset{iid}{\sim} f(x|\theta) = \theta e^{-\theta x}$, $\pi(\theta) = \beta e^{-\beta\theta}$, $\theta > 0$.

- May have more than one choice of the estimator of the parameter.
  - Need to evaluate the estimators so that we can choose the best one.
  - Need a criterion to evaluate the estimator. (Unbiasedness, MSE, Consistency, BLUE, UMVUE)

Definition

An estimator $W(X_1, \cdots, X_n)$ of a parametric function $\tau(\theta)$ is said to be an *unbiased* estimator (UE) if

$$E_\theta(W) = \tau(\theta), \quad \text{for all} \quad \theta \in \Theta.$$

Definition

The function

$$\mathrm{BIAS}_\theta(W) = E_\theta(W) - \tau(\theta)$$

is called the *bias* of W as an estimator of $\tau(\theta)$.

$\triangleright$ Example: $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Find an unbiased estimator for μ and σ^2. What about the MLE of σ^2?

Q) If W is an UE of θ, then is $\tau(W)$ UE of $\tau(\theta)$? Yes/No

$\triangleright$ Example-Continued: S^2 is an UE of σ^2. Is S unbiased for σ?

$E(S) - \sigma \approx \sigma/(4n)$.

Definition

The *Mean Squared Error (MSE)* of an estimator W of a parameter θ is the function of θ defined by

$$\mathrm{MSE}_\theta(W) = E_\theta(W - \theta)^2.$$

▷ Note: $\mathrm{MSE}_\theta(W) = Var_\theta(W) + [\mathrm{BIAS}_\theta(W)]^2$

▷ Example-continued: Which one is better,

$$\hat{\sigma}^2 = n^{-1}(n-1)S^2, \quad \tilde{\sigma}^2 = S^2?$$

(We compare MSEs of the two estimators. See Example 7.3.4.)

$\triangleright$ Example: $X_1, \cdots, X_n \overset{iid}{\sim}$ Bernoulli(p). With $Beta(\alpha, \beta)$ prior, the Bayes estimator is

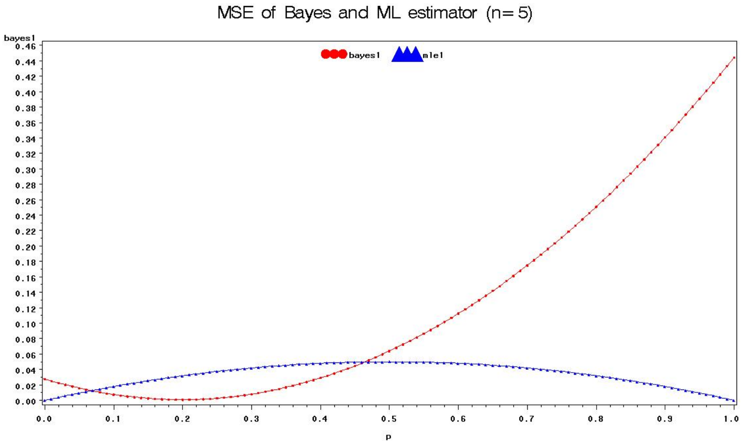
$$\text{Posterior Mean: } \hat{p}^{Bayes} = \frac{n\bar{X} + \alpha}{\alpha + \beta + n}.$$

The MLE of p is $\hat{p}^{MLE} = \bar{X}$. Compare the MSE of the two estimators.

See Example 7.3.5. for the details. Note the MSE of $\hat{p}^{Bayes}$ depends on α, β.

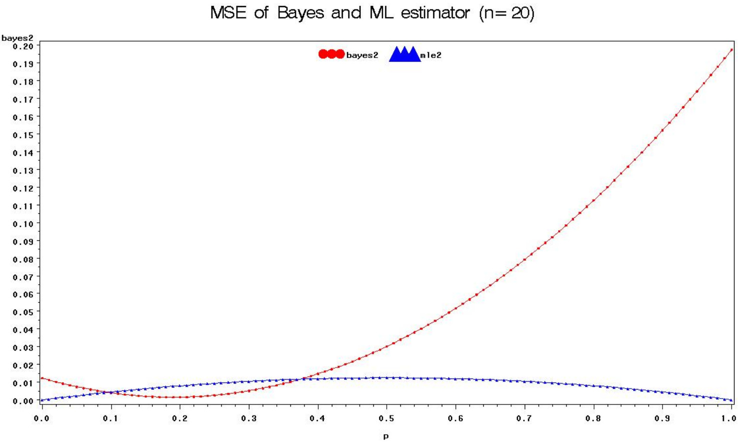
Ch 7. Point estimation

Mean Squared Error: MSE



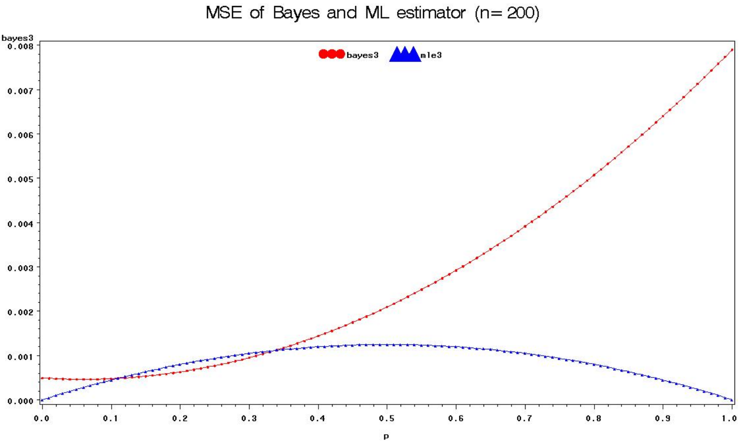
Ch 7. Point estimation

Mean Squared Error: MSE



Ch 7. Point estimation

Mean Squared Error: MSE



Definition

Let $f(\mathbf{x}|\theta) = f(x_1, \cdots, x_n|\theta)$ be the pdf/pmf of $X_1, \cdots, X_n$. An estimator W is said to be *Uniformly Minimum Variance Unbiased Estimator* (UMVUE) for $\tau(\theta)$ if

1. $E_\theta(W) = \tau(\theta)$
2. $Var_\theta(W) < \infty$
3. For any other UE of $\tau(\theta)$, say $\tilde{W}$,

$$Var_\theta(W) \leq Var_\theta(\tilde{W}) \ \text{ for all } \ \theta \in \Theta$$

▷ Note: UMVUE may not exist. If it does, it is essentially unique.

Then, how to get the UMVUE?

1. Using Cramér-Rao Lower Bound (CRLB)
2. Using complete sufficient statistic and Rao-Blackwell [Theorem 7.3.17], we apply Lehmann-Scheffé Theorem [Theorem 7.5.1]. (Also, see Theorem 7.3.23 where '*best*' means the minimum variance.)

- Idea of using CRLB;
  Show that for any UE, $\tilde{W}$, of $\tau(\theta)$,

$$Var_\theta(\tilde{W}) \geq c(\theta) \text{ for all } \theta \in \Theta$$

and if we can find an UE, W, such that

$$Var_\theta(W) = c(\theta) \text{ for all } \theta \in \Theta$$

then we can conclude W is the UMVUE of $\tau(\theta)$.

Theorem

Let $f(\mathbf{x}|\theta)$ be the pdf/pmf of $X_1, \cdots, X_n$. Assume

1. *Θ is an open space(subset) of R.*
2. *$\{\mathbf{x} : f(\mathbf{x}|\theta) > 0\}$ does not depend on θ.*
3. *$\partial f(\mathbf{x}|\theta)/\partial\theta$ exist on Θ*
4. *For any estimator $\tilde{W}$ with $E_\theta \tilde{W}^2 < \infty$, for all $\theta \in \Theta$, we have*

$$\frac{\partial}{\partial\theta} E_\theta \tilde{W} = \begin{cases} \int \tilde{W} \, [\frac{\partial}{\partial\theta} f(\mathbf{x}|\theta)] d\mathbf{x} \\ \sum \tilde{W} \, [\frac{\partial}{\partial\theta} f(\mathbf{x}|\theta)] \end{cases}$$

5. $$E_\theta \left[\left(\frac{\partial \ln f(\mathbf{x}|\theta)}{\partial\theta} \right)^2 \right] < \infty$$

Theorem (Continued)

Then for any UE of a differentiable parametric function $\tau(\theta)$,

$$Var_\theta(W) \geq \frac{[\tau(\theta)']^2}{E_\theta\left[\left(\frac{\partial \ln f(\mathbf{x}|\theta)}{\partial\theta}\right)^2\right]}.$$

▷ Note:

- The five conditions are called CR regularity conditions.
- $$I_n(\theta) = E_\theta\left[\left(\frac{\partial \ln f(\mathbf{x}|\theta)}{\partial\theta}\right)^2\right]$$

  is called Fisher information of the sample.

- When
$$\frac{\partial^2 \ln f(\mathbf{x}|\theta)}{\partial\theta^2}$$
exist, then by Lemma 7.3.11,
$$I_n(\theta) = -E_\theta\left[\left(\frac{\partial^2 \ln f(\mathbf{x}|\theta)}{\partial\theta^2}\right)\right]$$
- The five conditions are usually satisfied with exponential family.
- If $X_1, \cdots, X_n \overset{iid}{\sim} f(x|\theta)$ then
$$Var_\theta(W) \geq \frac{[\tau(\theta)']^2}{nE_\theta\left[\left(\frac{\partial \ln f(x|\theta)}{\partial\theta}\right)^2\right]}.$$

$\triangleright$ Example: $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, σ^2 in known. Find the CRLB and the UMVUE of μ.

$\triangleright$ Example: $X_1, \cdots, X_n \overset{iid}{\sim}$ Poisson(λ). Find the CRLB and the UMVUE of λ.

$\triangleright$ Example (We cannot apply CRLB here):
$X_1, \cdots, X_n \overset{iid}{\sim} Uniform(0, \theta)$. Then, by the factorization theorem, $X_{(n)}$ is a sufficient statistic. Example 6.2.23 shows $X_{(n)}$ is complete with $E(X_{(n)}) = \frac{n}{n+1}\theta$. Thus $\frac{n+1}{n}X_{(n)}$ is an UE for θ. Then, we apply Lehmann and Scheffé. (However, $\frac{n+1}{n}Var(X_{(n)}) = \frac{\theta^2}{n(n+2)} < \frac{\theta^2}{n}$ =CRLB in Example 7.3.13. One condition of CRLB is violated.) Note: we need completeness of $X_{(n)}$, not $X_i's$.

Theorem (Corollary 7.3.15 (Attainment))
$X_1, \cdots, X_n \overset{iid}{\sim} f(x|\theta)$. Let W be an unbiased estimator of $\tau(\theta)$. Then $Var_\theta(W)$ attains the CRLB if and only if

$$a(\theta)\,[W - \tau(\theta)] = \frac{\partial}{\partial\theta} \ln[f(x_1, \cdots, x_n|\theta)]$$

for some function $a(\theta)$.

▷ Example 7.3.16 : $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. (Here, $\tau(\theta) = \sigma^2$.)

▷ Example: $X_1, \cdots, X_n \overset{iid}{\sim}$ Exponential(λ). (Here, $\tau(\theta) = \lambda$.)

Theorem (Theorem 7.3.17 Rao-Blackwell)

Let W be any unbiased estimator of $\tau(\theta)$ and let T be a sufficient statistic for θ. Define $\phi(T) = E_\theta(W|T)$. Then

$$E_\theta[\phi(T)] = \tau(\theta)$$

and

$$Var_\theta[\phi(T)] \leq Var_\theta[W]$$

for all θ. That is, $\phi(T)$ is a uniformly better unbiased estimator of $\tau(\theta)$ than W.

Proof: See the proof of Theorem 7.3.17. It is very simple.

Theorem (Theorem 7.3.19)

If W is a best estimator of $\tau(\theta)$, then W is unique.

Theorem (Lehmann-Scheffe)

Let $X_1, \cdots, X_n$ have joint pmf/pdf $f(\mathbf{x}:\theta)$, $\theta \in \Theta$. Suppose T is a complete and sufficient statistic. If $\phi(T)$ is an unbiased estimator of $\tau(\theta)$ and it is a function of T only then $\phi(T)$ is the UMVUE of $\tau(\theta)$.

▷ Note:

1. If we can find an unbiased estimator $\phi(T)$ of $\tau(\theta)$ which is a function of CSS T only then it is the UMVUE.
2. For any unbiased estimator of $\tau(\theta)$, W, $E(W|T)$ is the UMVUE of $\tau(\theta)$.

$\triangleright$ Example: $X_1, \cdots, X_n \overset{iid}{\sim}$ Binomial(k, θ). Find the UMVUE of θ and $k\theta(1-\theta)^{k-1}$.

▷ Example: $X_1, \cdots, X_n \overset{iid}{\sim}$ Geometric(p). Find the UMVUE of p.

▷ Example: $X_1, \cdots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$. Find the UMVUE of θ^2.

- Data: $\mathbf{X}$
- Model(Distribution): $f(\mathbf{x}|\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$
- Action space (set of possible decisions): $\mathcal{A}$
  Point estimation: $\mathcal{A} = \Theta$
  Testing: $\mathcal{A}$={Reject H_0, Accept H_0}
- Loss function: $L(\boldsymbol{\theta}, a)$, a is an action.
- Decision rule: $\delta(\mathbf{x})$: Sample space $\rightarrow \mathcal{A}$
- Risk function: Expected loss

$$R(\boldsymbol{\theta}, \delta) = E\left[L(\boldsymbol{\theta}, \delta(\mathbf{X}))\right] = \int L(\boldsymbol{\theta}, \delta(\mathbf{x})) f(\mathbf{x}|\boldsymbol{\theta}) dx$$

- Goal: Find $\delta(\mathbf{x})$ that has small risk somehow.

Definition

A real valued function $L(\boldsymbol{\theta}, a)$ satisfying

1. $L(\boldsymbol{\theta}, a) \geq 0$ for all θ, a
2. $L(\boldsymbol{\theta}, a) = 0$ for $a = \theta$

is called a *loss function* of the action a.

Definition

Let $\delta(\mathbf{X})$ be an estimator of a parametric function $\tau(\boldsymbol{\theta})$. Then

$$R(\boldsymbol{\theta}, \delta) = E\left[L(\boldsymbol{\theta}, \delta(\mathbf{X}))\right]$$

is called the *risk function* of $\delta(\mathbf{X})$ in estimating $\tau(\boldsymbol{\theta})$.

- Example:

1. Squared error loss

$$L(\theta, \delta(\mathbf{X})) = (\delta(\mathbf{X}) - \theta)^2,\ R(\boldsymbol{\theta}, \delta) = E\left[(\delta(\mathbf{X}) - \theta)^2\right]$$

2. Absolute error loss

$$L(\theta, \delta(\mathbf{X})) = |\delta(\mathbf{X}) - \theta|,\ R(\boldsymbol{\theta}, \delta) = E\left[|\delta(\mathbf{X}) - \theta|\right]$$

3. Stein's loss

$$L(\theta, \delta(\mathbf{X})) = \frac{\delta(\mathbf{X})}{\theta} - 1 - \ln\left(\frac{\delta(\mathbf{X})}{\theta}\right)$$

Definition

1. An estimator $\delta_1(\mathbf{X})$ is said to be at least *as good as* another estimator $\delta_2(\mathbf{X})$ if
$$R(\theta, \delta_1(\mathbf{X})) \leq R(\theta, \delta_2(\mathbf{X}))$$
for all $\theta \in \Theta$.
2. An estimator $\delta_1(\mathbf{X})$ is *better than* $\delta_2(\mathbf{X})$ if
$$R(\theta, \delta_1(\mathbf{X})) \leq R(\theta, \delta_2(\mathbf{X}))$$
for all $\theta \in \Theta$ and
$$R(\theta, \delta_1(\mathbf{X})) < R(\theta, \delta_2(\mathbf{X}))$$
for at least one $\theta \in \Theta$.

Definition

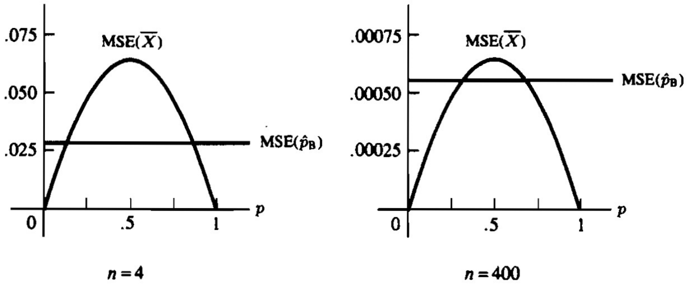
An estimator $\delta(\mathbf{X})$ is said to be *admissible* if no other estimator is better than $\delta(\mathbf{X})$.

$\triangleright$ Example: $X_1, \cdots, X_n \overset{iid}{\sim} N(\theta, 1)$. $L(\theta, \delta(\mathbf{X})) = (\delta(\mathbf{X}) - \theta)^2$.
$\delta_1(\mathbf{X}) = c$, $\delta_2(\mathbf{X}) = \bar{X}$.

▷ Example: $X_1, \cdots, X_n \overset{iid}{\sim}$ Bernoulli(θ). $L(\theta, \delta(\mathbf{X})) = (\delta(\mathbf{X}) - \theta)^2$.
$\delta_1(\mathbf{X}) = \bar{X}$,

$$\delta_2(\mathbf{X}) = \frac{n\bar{X} + \sqrt{n/4}}{n + \sqrt{n}}.$$

$\delta_2(\mathbf{X})$ is a Bayes estimator with a *Beta* prior. See Example 7.3.5. for the details.



Ch 7. Point estimation

Decision Theory: Loss function optimality

$\triangleright$ Example: $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Estimation of σ^2 using different loss function. We restrict our estimator of the form $\delta_b(\mathbf{X}) = bS^2$. The loss function considered are squared error loss, Stein's loss.

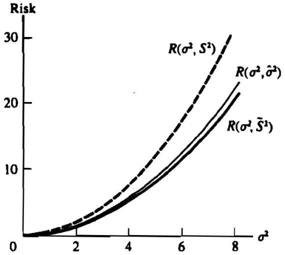


Figure 7.3.2. *Risk functions for three variance estimators in Example 7.3.26*

- In general, there does not exist an estimator $\delta(\mathbf{X})$ such that for any other estimator $\tilde{\delta}(\mathbf{X})$ we have

$$R(\theta, \delta(\mathbf{X})) \leq R(\theta, \tilde{\delta}_2(\mathbf{X}))$$

  for all $\theta \in \Theta$.
- To define the best estimator w.r.t. the given loss function, we can proceed two ways.
  1. Restrict attention to a smaller class of estimators such as unbiased estimators or linear estimators
  2. Define a criterion for comparing estimators such minimax or Bayes rule

Definition

An estimator $\delta(\mathbf{X})$ is called a *minimax estimator* if

$$\max_{\theta\in\Theta} R(\theta, \delta(\mathbf{X})) \leq \max_{\theta\in\Theta} R(\theta, \tilde{\delta}(\mathbf{X}))$$

for all other estimator $\tilde{\delta}(\mathbf{X})$.

Definition

The *Bayes risk* of an estimator $\delta(\mathbf{X})$ w.r.t. prior distribution $\pi(\theta)$ is defined as

$$B(\pi, \delta(\mathbf{X})) = E_\pi\left[R(\theta, \delta(\mathbf{X}))\right] = \int R(\theta, \delta(\mathbf{X}))\pi(\theta)d\theta$$

Definition

An estimator $\delta(\mathbf{X})^{\pi}$ is said to be a *Bayes estimator* w.r.t. prior distribution $\pi(\theta)$ if it minimizes Bayes risk over all estimators. That is

$$B(\pi, \delta(\mathbf{X})^{\pi}) = \inf_{\tilde{\delta}} B(\pi, \tilde{\delta}(\mathbf{X}))$$

Theorem

Consider a point estimation problem for a real-valued parameter θ. The Bayes estimator is $E(\theta|\mathbf{X})$ for squared error loss and median of $\pi(\theta|\mathbf{X})$ for absolute error loss.

▷ Example 7.3.30: $X_1, \cdots, X_n \overset{iid}{\sim}$ Bernoulli(p). $p \sim$ Uniform$(0,1)$. Find the Bayes estimator with squared error loss function.

Table 7.3.1. *Three estimators for a binomial p*

$n = 10$	prior $\pi(p) \sim$ uniform$(0,1)$		
y	MLE	Bayes absolute error	Bayes squared error
0	.0000	.0611	.0833
1	.1000	.1480	.1667
2	.2000	.2358	.2500
3	.3000	.3238	.3333
4	.4000	.4119	.4167
5	.5000	.5000	.5000
6	.6000	.5881	.5833
7	.7000	.6762	.6667
8	.8000	.7642	.7500
9	.9000	.8520	.8333
10	1.0000	.9389	.9137