STAT 5430
Lecture 11, F, Feb 14

- Homework 3 is assigned in Canvas (due by Monday, Feb 17 by midnight) (practice on CRLB/UMVUE/Bayes)
- Solutions to Homeworks 1-2 posted.
- Exam 1 is scheduled for W, Feb 26 (two weeks) 6:15-8:15 PM (Sned seminar room)
  - No regular class on W, Feb 26
  - See Canvas for study guide, practice exams
  - can bring 1 page formula sheet (front/back) with anything on it
  - see Canvas for a "canned" sheet
  - I'll provide table with STAT 5420 distributions (see Canvas)

ASIDE/
SKIP

Elements of Decision Theory

A Note on Bayes and Minimax Estimators

As we have seen, the minimax and Bayes principles are ways of determining estimators based on risk function considerations. There are also some interesting connections between the two types of estimators. As as example, the following result shows how to find a minimax estimator (which is hard) from a Bayes estimator (which can be done more easily).

Theorem: For some loss function $L(t,\theta)$, if T^* is a Bayes estimator with respect to *some* prior and the risk of T^* is constant (i.e., $R_{T^*}(\theta) = c$ for all $\theta \in \Theta$), then T^* is the minimax estimator under the same loss function.[1]

.

Example: For $X_1, \ldots, X_n$ iid Bernoulli(θ), $\theta \in \Theta = (0,1)$, find the minimax estimator of θ under the loss function $L(t,\theta) = (\theta - t)^2/\{\theta(1-\theta)\}$.

Solution: For this loss function, the Bayes estimator of θ with respect to the uniform(0,1) prior on Θ is given by $T_0 = \bar{X}_n$ (see Homework 3). Also note that, for any $\theta \in (0,1)$,

$E_\theta X_1 = \theta = E_\theta \bar{X}_n$

$\mathrm{Var}_\theta(X_1)/n = \dfrac{\theta(1-\theta)}{n}$

$$\begin{aligned} R_{T_0}(\theta) &= \mathrm{E}_\theta\left(\frac{(\theta - \bar{X}_n)^2}{\theta(1-\theta)}\right) = \frac{1}{\theta(1-\theta)}\mathrm{MSE}_\theta(\bar{X}_n) \\ &= \frac{1}{\theta(1-\theta)}\mathrm{Var}_\theta(\bar{X}_n) \qquad \text{since } \bar{X}_n \text{ is unbiased} \\ &= \frac{1}{n} \qquad \text{since } \mathrm{Var}_\theta(X_1) = \theta(1-\theta) = n\mathrm{Var}_\theta(\bar{X}_n) \end{aligned}$$

risk of T_0

constant

Hence, T_0 has constant risk so that, by the theorem, $T_0 = \bar{X}_n$ is also the minimax estimator of θ.

[1] *Proof*: Suppose that T^* is a Bayes estimator with respect to a prior pdf $\pi(\theta)$ on Θ. Then the Bayes risk of T^* is $BR_{T^*} = \int_\Theta R_{T^*}(\theta)\pi(\theta)d\theta = c\int_\Theta \pi(\theta)d\theta = c$, using that $R_{T^*}(\theta) = c$ is constant and $\int_\Theta \pi(\theta)d\theta = 1$. Since T^* is the Bayes estimator, it has minimal Bayes risk by definition so that, for any other estimator T, we have

$$c = BR_{T^*} \le BR_T = \int_\Theta R_T(\theta)\pi(\theta)d\theta \le \int_\Theta \left[\max_{\theta\in\Theta} R_T(\theta)\right]\pi(\theta)d\theta = \left[\max_{\theta\in\Theta} R_T(\theta)\right]\int_\Theta \pi(\theta)d\theta = \max_{\theta\in\Theta} R_T(\theta)$$

or $c \le \max_{\theta\in\Theta} R_T(\theta)$ for any estimator T. Because $\max_{\theta\in\Theta} R_{T^*}(\theta) = c$, it must be that T^* is minimax.

STAT 5430: Summary to date
Where we have been
& where we are headed

- Completed
  - Introduction to Statistical Inference
  - Point Estimation
    - MME/MLE
  - Criteria for Evaluating Point Estimators
    - bias, variance, UMVUE, MSE
  - Elements of Decision Theory
    - General concepts (e.g., risk)
    - Minimax & Bayes principles
    - Finding Bayes Estimators

- Next: Large Sample Properties of Estimators
  - General concepts: consistency, MSE consistency, asymptotic unbiased
  - Showing consistency
  - Asymptotic efficiency
  - Asymptotic normality of MLEs
  - Delta method

Large Sample Properties of Estimators

Asymptotic

Terminology

$n \equiv$ "sample size"

Definition: Let $\{T_n\}$ be a sequence of estimators of a parametric function $\gamma(\theta)$.

1. $\{T_n\}$ is called **consistent** for $\gamma(\theta)$ if, for any $\epsilon > 0$, "little distance"

   (a) $\lim_{n\to\infty} P_\theta\left(|T_n - \gamma(\theta)| < \epsilon\right) = 1$, for all $\theta \in \Theta$,

   (b) or, equivalently, if $\lim_{n\to\infty} P_\theta\left(|T_n - \gamma(\theta)| \geq \epsilon\right) = 0$, for all $\theta \in \Theta$.

   (c) or, equivalently, if "T_n converges in probability to $\gamma(\theta)$ as $n \to \infty$" i.e, $T_n \xrightarrow{p} \gamma(\theta)$ as $n \to \infty$.

2. $\{T_n\}$ is called **mean squared error consistent (MSEC)** for $\gamma(\theta)$ if

$$\lim_{n\to\infty} \mathrm{MSE}_\theta(T_n) \equiv \lim_{n\to\infty} \mathrm{E}_\theta\left(\left[T_n - \gamma(\theta)\right]^2\right) = 0, \quad \forall\theta \in \Theta.$$

3. $\{T_n\}$ is called **asymptotically unbiased** for $\gamma(\theta)$ if

$$\lim_{n\to\infty} \mathrm{E}_\theta(T_n) = \gamma(\theta), \quad \forall\theta \in \Theta.$$

$$\lim_{n\to\infty} [\underbrace{E_\theta(T_n) - \gamma(\theta)}_{\equiv \text{bias } b_\theta[T_n]}] = 0$$

Definitions 1-3 represent different senses in which T_n can be "close" to $\gamma(\theta)$ as sample size $n \to \infty$

Large Sample Properties of Estimators

Terminology, cont'd

Example. Let $X_1, \dots, X_n$ be iid with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$.

parameters $\theta = (\mu, \sigma^2)$

- Consider the sequence of estimators $T_n = \bar{X}_n$, $n \geq 1$, of μ.

It turns out $\bar{X}_n \xrightarrow{P} EX_1 = \mu$ as $n \to \infty$ by WLLN (later)

or pick $\epsilon > 0$,

$$P_\theta(|\bar{X}_n - \mu| \geq \epsilon) = P_\theta(|\bar{X}_n - \mu|^2 \geq \epsilon^2)$$

Markov inequality

$$\leq \frac{E_\theta(|\bar{X}_n - \mu|^2)}{\epsilon^2} = \frac{Var_\theta(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

$\to 0$ as $n \to \infty$ for any $\theta = (\mu, \sigma^2)$

$\therefore \bar{X}_n$ is consistent for μ!

- $\text{MSE}_\theta(\bar{X}_n) = \text{E}_\theta\left(\left[\bar{X}_n - \mu\right]^2\right)$

$= Var_\theta(\bar{X}_n) = \frac{\sigma^2}{n} \to 0$ as $n \to \infty$ for any $\theta = (\mu, \sigma^2)$

$\therefore \bar{X}_n$ is MSEC for μ too!

$\Leftarrow$ MLE/MOM of σ^2 when data are normal

- Consider the sequence of estimators $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2 = \frac{n-1}{n}S^2$ of σ^2, where $S^2 = \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2$, $n \geq 1$.

$$E_\theta(\hat{\sigma}_n^2) = \frac{n-1}{n} E_\theta(S^2) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

So $\hat{\sigma}_n^2$ is biased, but $\lim_{n \to \infty} E_\theta(\hat{\sigma}_n^2) = \sigma^2$.

Hence, $\hat{\sigma}_n^2$ is asym. unbiased.

Large Sample Properties of Estimators

Tools for Showing Consistency

Question: Is $\hat{\sigma}_n^2$ consistent for σ^2?

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - (\bar{X}_n)^2$$

notice 2 types of sample means (for X_i^2 & X_i)

Two useful results for establishing consistency $W_i \equiv \begin{pmatrix} W_{i,1} \\ \vdots \\ W_{i,k} \end{pmatrix}$

1. (WLLN): If $W_1, W_2, \ldots$ are iid k-dimensional random vectors with $\mathrm{E}|W_1| \equiv \mathrm{E}|W_{1,1}| + \cdots + \mathrm{E}|W_{1,k}| < \infty$ where $W_1 = (W_{1,1}, \ldots, W_{1,k})'$, then

weak law of large numbers

$$\bar{W}_n = \frac{1}{n}\sum_{i=1}^n W_i \xrightarrow{p} \mathrm{E}(W_1) \equiv \begin{pmatrix} \mathrm{E}(W_{1,1}) \\ \vdots \\ \mathrm{E}(W_{1,k}) \end{pmatrix} \quad \text{as } n \to \infty.$$

k sample averages

each component of W_1 has mean

$|W_i| = |W_{i,1}| + \cdots + |W_{i,k}|$

2. (Continuous Mapping Theorem): Suppose $Y_n \xrightarrow{p} Y$, where $\{Y_n\}, Y$ are k-dimensional random vectors, $k \geq 1$. Then,

   (a) for any continuous function $g : \mathbb{R}^k \to \mathbb{R}^p$ $(p \geq 1)$, we have that

   $$g(Y_n) \xrightarrow{p} g(Y) \quad \text{as } n \to \infty.$$

   special case of (a)

   $Y_n \xrightarrow{p} c$ as $n \to \infty$

   (b) if $P(Y = c) = 1$ for some $c \in \mathbb{R}^k$ (i.e., Y is essentially a constant c) and $g : \mathbb{R}^k \to \mathbb{R}^p$ is *continuous at* $c \in \mathbb{R}^k$, then

   $$g(Y_n) \xrightarrow{p} g(c) \quad \text{as } n \to \infty.$$

Large Sample Properties of Estimators

Showing Consistency

Example. Let $X_1, \ldots, X_n$ be iid with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$. Show $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X}_n)^2$ is consistent for σ^2.

Let $W_i = \begin{pmatrix} X_i^2 \\ X_i \end{pmatrix}$, for $i \geq 1$.

Then $W_1, \ldots, W_n$ are iid random vectors

with $E|W_1| = E|X_1^2| + E|X_1| < \infty$.

By WLLN, $\bar{W}_n = \frac{1}{n}\sum_{i=1}^{n} W_i = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} X_i^2 \\ \frac{1}{n}\sum_{i=1}^{n} X_i \end{pmatrix} \xrightarrow{P} \begin{pmatrix} EX_1^2 \\ EX_1 \end{pmatrix}$

as $n \to \infty$.

Note $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X}_n)^2$

$= g\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2, \frac{1}{n}\sum_{i=1}^{n} X_i\right)$, where

$g(x,y) = x - y^2$

Since $g(x,y)$ is continuous on $\mathbb{R}^2$,

by continuous mapping theorem,

$\hat{\sigma}_n^2 = g\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2, \frac{1}{n}\sum_{i=1}^{n} X_i\right) \xrightarrow{P} g(EX_1^2, EX_1)$

$= EX_1^2 - (EX_1)^2$

$= \sigma^2$.

$\therefore$ $\hat{\sigma}_n^2$ is consistent for σ^2 //