Math 502AB - Lecture 23

Dr. Jamshidian

November 15, 2017

1 Lecture - Part 1

1.0.1 Theorem:

Let $X_1, ..., X_n \sim f(X|\theta)$ (**iid**), where $f(X|\theta)$ satisfies the conditions of the *Cramer-Rao Lower Bound*. Let:

$$\mathcal{L}(\theta|X) = \prod_{i=1}^{n} f(x_i|\theta)$$

If $W(X)$ is any unbiased estimator of $\tau(\theta)$, then $W(X)$ attains the *CRLB* **iff** there exists a function $a(\theta)$ such that:

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

Examples:

1. Let $X_1, ..., X_n \sim Gamma\left(1, \frac{1}{\theta}\right)$ (**iid**). Investigate *UMVUE* for θ.

$$\mathcal{L}(\theta|X_1, ..., X_n) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum x_i}$$

So what might be a sufficient statistic? We have, by the factorization theorem, $\sum_{i=1}^{n} x_i$ is a sufficient statistic.

$$\ell(\theta) = n\log(\theta) - \theta \sum x_i \Rightarrow \hat{\theta} = \frac{1}{\bar{X}} = \frac{n}{\sum_{i=1}^{n} x_i}$$

Is $\hat{\theta}$ *unbiased??*

$$\sum x_i \sim Gamma\left(n, \frac{1}{\theta}\right)$$
$$E(\hat{\theta}) = nE\left(\frac{1}{\sum x_i}\right) = n\int_0^\infty \frac{1}{y} \cdot \frac{\theta^n}{\Gamma(n)} y^{n-1} e^{-\theta y} dy = \frac{n\theta}{n-1}$$

So, no it isn't. But if we let $W(X) = \frac{n-1}{n}\hat{\theta}$, then $W(X)$ *is* unbiased. **Does $W(X)$ obtain the CRLB?** Using the theorem above, we have:

$$\begin{aligned}
\frac{\partial}{\partial\theta}\log\mathcal{L}(\theta|X) &= \frac{\partial}{\partial\theta}\left[n\log\theta - \theta\sum x_i\right] \\
&= \frac{n}{\theta} - \sum x_i = \frac{n-\theta\sum x_i}{\theta} \quad \textbf{(1)} \\
w(X)-\theta &= \frac{n-1}{\sum x_i} - \theta = \frac{n-1-\theta\sum x_i}{\sum x_i} \quad \textbf{(2)}
\end{aligned}$$

Since $\frac{\textbf{(1)}}{\textbf{(2)}}$ is **not** a function of θ only, $W(X)$ does **not** achieve the *CRLB*.

So, we want to know: is $W(X)$ the *best* unbiased estimator?

1.0.2 Theorem (7.3.23)

Let T be a complete sufficient statistic for a parameter θ, and let $g(T)$ be an estimator based only on T, such that $E[g(T)] = \tau(\theta)$. Then $g(T)$ is the *best unique unbiased estimator* of $\tau(\theta)$

1.0.3 The Rao-Blackwell Theorem

Given two random variables X and Y

$$\begin{aligned}
E[X] &= E[E(X|Y)] \\
Var(X) &= Var(E[X|Y]) + E[Var(X|Y)] \geq Var(E[X|Y])
\end{aligned}$$

Now, let W be a statistic for estimating θ, and let T be a sufficient statistic. Let $E[W] = \theta$, and define $\phi(T) = E[W|T]$.

1. $\phi(T)$ is independent of θ, because T is a sufficient statistic. Thus $\phi(T)$ is itself an estimator of θ.
2. $E[\phi(T)] = E[E(W|T)] = E[W] = \theta$
3. $Var(\phi(T)) = Var\left(E[W|T]\right) \leq Var(W)$

1.0.4 Theorem:

Let $X_1, ..., X_n \sum f(X,\theta)$ (**iid**). Let $Y = U(X_1, ..., X_n)$ be a sufficient statistic for θ, provided it exists. If $\hat{\theta}$ is the unique MLE of θ, it must be a function of Y.

Proof:

Consider the likelihood function

$$
\begin{aligned}
L(\theta|X_1, ..., X_n) &= \prod_{i=1}^{n} f(x_i|\theta) \\
\text{By factorization theorem: } &= f_Y[u(x_1, ..., x_n), \theta] H(x_1, ..., x_n)
\end{aligned}
$$

1.1 Loss Functions

Sometimes we are interested in obtaining estimators that achieve certain goals. In such cases, we define "loss functions" to achieve our goal.

Examples:

1. Suppose that a is to estimate θ. Two simple loss functions we can consider are:
   (a) Absolute error loss $= |\theta - a|$
   (b) Squared error loss $= (\theta - a)^2$
2. The following loss function penalized *overestimation:*

$$
L(\theta, a) = \begin{cases} (a-\theta)^2 & a < \theta \\ 10(a-\theta)^2 & a > \theta \end{cases}
$$

1.1.1 Risk Function

Suppose that $\delta(X)$ is an estimator of θ. Then the **risk function** is defined by:

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

Therefore, the risk is the *average loss at a given value* of θ.

Example:

$$
\begin{aligned}
L(\theta, \delta(X)) &= (\theta - \delta(X))^2 \\
\mathcal{R}(\theta, \delta(X)) &= E[\theta - \delta(X)]^2 = MSE_\theta(\delta(X))
\end{aligned}
$$

Example:

Let $X_1, ..., X_n \sim N(\mu, \sigma^2)$. Consider estimating σ^2, using squared error loss. Let's consider estimators of the form $\delta_b(X) = bS^2$, where $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2$.

Recall that $E(S^2) = \sigma^2$ and $Var(S^2) = \frac{2\sigma^4}{n-1}$. Then, lets look at the risk:

$$\begin{aligned}
\mathcal{R}(\theta, \delta(X)) &= Var(bS^2) + \left[E(bS^2) - \sigma^2\right]^2 \\
&= \frac{b^2(2\sigma^4)}{n-1} + [b\sigma^2 - \sigma^2] \\
&= \left[\frac{2b^2}{n-1} + (b-1)^2\right]\sigma^4
\end{aligned}$$

Now, let's consider three different b's. For S^2, we have $b^2 = 1$, and for $\hat{\sigma}^2$, we have $b^2 = \frac{n-1}{n}$. But what about the **optimal** b?

$$\begin{aligned}
g(b) &= \frac{2b^2}{n-1} + (b-1)^2 \\
g'(b) &= \frac{4b}{n-1} + 2(b-1) = 0 \\
\Rightarrow b &= \frac{n-1}{n+1}
\end{aligned}$$

1.1.2 Stein's Loss Function

$$L(\sigma^2, a) = \frac{a}{\sigma^2} - 1 - \log\frac{a}{\sigma^2}$$

Let's consider the same case with this new loss function:

$$\begin{aligned}
\mathcal{R}(\sigma^2 bS^2) &= E\left[\frac{bS^2}{\sigma^2} - 1 - \log\frac{bS^2}{\sigma^2}\right] \\
&= bE\left[\frac{S^2}{\sigma^2}\right] - 1 - \log b - E\left[\log\frac{S^2}{\sigma^2}\right] \\
\Rightarrow g(b) &= b - 1 - \log b \\
g'(b) &= 1 - \frac{1}{b} = 0 \Rightarrow b = 1
\end{aligned}$$

1.2 Bayesian Point Estimation

In *Bayesian estimation*, we minimize what is referred to as the **Bayes Risk**:

$$\min_{\delta(X)} \int_{\Theta} \mathcal{R}(\theta, \delta(X))\pi(\theta)d\theta$$

Where $\pi(\theta)$ is the *prior*. It can actually be shown that this is equivalent to:

$$\min \int_{-\infty}^{\infty} L(\theta, \delta(X)) f_{\Theta|X}(\theta|x)\theta d\theta = \min E\left[L(\Theta, \delta(X))|X = x\right]$$

Then, if the loss function is:

$$\begin{aligned}
L(\theta, \delta(X)) &= (\theta - \delta(X))^2 \Rightarrow \delta(X) = E[\Theta|X] \\
L(\theta, \delta(X)) &= |\theta - \delta(X)| \Rightarrow \delta(X) = median(\theta|X)
\end{aligned}$$

Example:

Suppose $X_1, ..., X_n \sim Bernouli(\theta)$ (**iid**). We have:

$$\begin{aligned} \text{Prior:} \quad & \Theta \sim Beta(\alpha, \beta) \quad \text{Where } \alpha \text{ and } \beta \text{ are known} \\ \text{Posterior:} \quad & \Theta | X \sim Beta\left((\alpha + \sum x_i, \beta + n - \sum x_i\right) \end{aligned}$$

Thus, using the *squared error loss*, the Bayes estimate is:

$$\delta(X) = E[\Theta|X] = \frac{\alpha + \sum x_i}{\alpha + \beta + n}$$

Note, that we can rewrite $E(\Theta|X)$ as follows:

$$E[\Theta|X] = \left(\frac{n}{\alpha + \beta + n}\right) \frac{\sum x_i}{n} + \left(\frac{\alpha + \beta}{\alpha + \beta + n}\right) \frac{\alpha}{\alpha + \beta}$$

Thus, it is some linear combination of the mean of the prior and the maximum likelihood estimate.

2 Lecture - Part II

2.1 Chapter 8: Test of Hypotheses

We wish to decide between:

$$\begin{cases} H_0 : & \theta \in \Theta_0 \\ H_a : & \theta \in \Theta_a \end{cases}$$

Given data $X_1, ..., X_n$.

Example:

Suppose that we have two coins:

$$\begin{cases} \text{Coin 0:} & P(heads) = 0.5 \\ \text{Coin 1:} & P(heads) = 0.7 \end{cases}$$

I choose a coin and flip it 5 times, and tell you $X =$ the number of heads. But based on this information, you are to decide which coin I flipped:

$$\begin{cases} H_0 : & p = 0.5 \\ H_a : & p = 0.3 \end{cases}$$

	Coin 1	Coin 2
X	$P_0(X)$	$P_1(X)$
0	0.031	0.0024
1	0.156	0.0028
2	0.313	0.132
3	0.313	0.309
4	0.156	0.36
5	0.031	0.168

Suppose that I tell you that I got 1 head:

$$\frac{P_0(1)}{P_1(1)} = \frac{0.156}{0.12}$$

X	$=$	0	1	2	3	4	5
$\Rightarrow \frac{P_0(X)}{P_1(X)}$	$=$	12.9	5.5	2.4	1.01	0.43	0.18

The values greater than 1 lead us to accept H_0 (ie. the acceptance region).

2.1.1 Possible Errors:

- **Type 1 Error:** Rejecting H_0 when it is true

$$P(X = 4, 5|p = 0.5) = 0.1875$$

- **Type 2 Error:** Failing to reject H_0 when it is false

$$P(X = 0, 1, 2, 3|p = 0.7) = 0.47$$

Consider another decision rule:

- Accept H_0 if: $X = \{0, 1, 2, 3, 4\}$
- Accept H_1 if: $X = \{5\}$
  - Type 1 Error: $P(X = 5|p = 0.5) = 0.031$
  - Type 2 Error: $P(X = 0, 1, 2, 3, 4|p = 0.7) = 0.83$

So which do you choose? A common heuristic is to choose which one has a less serious Type 1 Error.

2.2 Likelihood Ratio Test

A method for construction rules for a test of hypothesis is called the *likelihood ratio test* (**LRT**).

Suppose that $X_1, ..., X_n \sim f(X|\theta)$ (**iid**). Then, the likelihood function:

$$\mathcal{L}(\theta|X) = \prod_{i=1}^{n} f(X_i|\theta)$$

Definition

The **likelihood ratio statistic** for testing:

$$\begin{cases} H_0: & \theta \in \Theta_0 \\ H_a: & \theta \in \Theta_0^c \end{cases}$$

Is:

$$\Lambda(X) = \frac{\max_{\theta \in \Theta_0} \mathcal{L}(\theta|X)}{\max_{\theta \in \Theta} \mathcal{L}(\theta|X)}$$

Where:

$$\Theta = \Theta_0 \cup \Theta_0^c$$

A likelihood ratio tst is any test that has a rejection region of the form:

$$\{X : \Lambda(X) \leq c\}$$

Where c is any constant satisfying $0 \leq c \leq 1$

Example:

Suppose $X_1, ..., X_n$ (**iid**) coming from:

$$f(X|\theta) = \frac{1}{\theta} e^{-x/\theta} \quad \theta > 0, x > 0$$

Construct a likelihood ratio test to test:

$$\begin{cases} H_0: & \theta = \theta_0 \\ H_a: & \theta \neq \theta_0 \end{cases}$$

$$\Theta_0 = \{\theta : \quad \theta = \theta_0\}$$
$$\Theta = \{\theta : \quad \theta > 0\}$$

We have:

$$\mathcal{L}(\theta|X) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-x/\theta} = \frac{1}{\theta^n} e^{-\frac{1}{\theta} \sum x_i}$$

Then,

$$\max_{\theta \in \Theta_0} \mathcal{L}(\theta|X) = \frac{1}{\theta^n} e^{-\frac{1}{\theta_0} \sum x_i} = \frac{1}{\theta_0^n} e^{-\frac{n\overline{X}}{\theta_0}}$$
$$\max_{\theta \in \Theta} \mathcal{L}(\theta|X) = \left(\frac{1}{\overline{X}}\right)^n e^{-n} \quad \text{(Recall that } \hat{\theta} = \overline{X}\text{)}$$

This gives our likelihood ratio as:

$$\Lambda(X) = \left(\frac{\overline{X}}{\theta_0}\right)^n e^n e^{-\frac{n\overline{X}}{\theta_0}}$$

Our rejection region is then of the form:

$$\left\{X : \left(\frac{\overline{X}}{\theta_0}\right)^n e^n e^{-\frac{n\overline{X}}{\theta_0}} < c\right\}$$
$$\left\{X : \left(\frac{\overline{X}}{\theta_0}\right)^n e^{-\frac{n\overline{X}}{\theta_0}} < c'\right\}$$

Let $t = \frac{\overline{X}}{\theta_0}$, and consider the function:

$$\begin{aligned} g(t) &= t^n \exp(-nt) \\ g'(t) &= t^{n-1} e^{-nt}(n - nt) = 0 \\ t &= 1 \quad g''(1) < 0 \end{aligned}$$

Thus, we have:

$$\Lambda(X) = \left\{X : \frac{\overline{X}}{\theta_0} < c_1 \text{ or } \frac{\overline{X}}{\theta_0} > c_2\right\}$$

Now, provided that H_0 is true (ie. $\theta = \theta_0$):

$$\frac{2X_i}{\theta_0} \sim \chi^2_{(2)} \Rightarrow \frac{2}{\theta_0}\sum_{i=1}^n x_i \sim \chi^2_{(2n)}$$

$$\frac{\overline{X}}{\theta_0} < c_1 \iff \frac{\sum x_i}{n\theta_0} < c_1 \iff \frac{2\sum x_i}{\theta_0} < 2c_1 n$$

Thus, the rejection region can be written as:

$$\Lambda(X) = \left\{X : \frac{2\sum x_i}{\theta_0} < c_1' \text{ or } \frac{2\sum x_i}{\theta_0} > c_2'\right\}$$

Suppose you are interested to test at $\alpha = 0.05$:

$$\begin{aligned} 0.05 &= P(\text{rejecting } H_0 \text{ — } H_0 \text{ is true}) \\ &= P\left(\chi^2_{(2n)} < c_1 \text{ or } \chi^2_{(2n)} > c_2\right) \end{aligned}$$