# My white board writing from week 6

1) I started with the Poisson($\theta$) example. $f(x,\theta)=\frac{e^{-\theta}\theta^x}{x!}, x=0,1,2,\dots$

$\mathbb{X}=(X_1,X_2,\dots,X_n)$ i.i.d Poisson($\theta$):

$L(\mathbb{X},\theta)=\frac{e^{-n\theta}\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$, hence $V(X,\theta)=\frac{\partial}{\partial\theta}\log L=-n+\frac{\sum_{i=1}^n x_i}{\theta}$

I considered two cases for a parameter $\tau(\theta)$ of interest:

a) $\tau(\theta)=\theta$. In this case $V(X,\theta)=\frac{n}{\theta}(\bar{X}-\theta)$ represents a factorization in which here we have a statistic (i.e. transformation of the data only, no parameter involved). Hence the CR bound is attainable and the statistic $\bar{X}$ attains it.

We can check also directly the attainability in this case:

$-\frac{\partial^2}{\partial\theta^2}\log L=\frac{\sum_{i=1}^n x_i}{\theta^2}$ implies $I_X(\theta)=E\left(-\frac{\partial^2}{\partial\theta^2}\log L\right)=\frac{n\theta}{\theta^2}=\frac{n}{\theta}$

and the CR bound for variance of an unbiased estimator of $\tau(\theta)=\theta$ is $\frac{(\tau'(\theta))^2}{I_X(\theta)}=\frac{1}{n/\theta}=\boxed{\frac{\theta}{n}}$

And a direct check shows: $Var(\bar{X})=\frac{1}{n^2}Var\left(\sum_{i=1}^n X_i\right)=\frac{\cancel{n}\theta}{n^{\cancel{2}}}$

b) $\tau(\theta)=e^{-\theta}=P(X_1=0)$. Then $V(X,\theta)=ne^{\theta}\left(\frac{1}{\theta}e^{-\theta}\bar{X}-e^{-\theta}\right)$

Since the quantity in the factorization now does depend on $\theta$ (i.e., is not a statistic), the CR Bound is not attainable by any unbiased estimator of $\tau(\theta)=e^{-\theta}$

HOWEVER, as we will see a bit later, $\left(1-\frac{1}{n}\right)^{n\bar{X}}$ is an unbiased estimator of $\tau(\theta)=e^{-\theta}$ that is the UMVUE (just that its variance, even if the smallest possible, is > then the Bound.

2) I then went through the proof of the Rao Blackwell theorem:

i) $E\hat{\tau}(T) = E_T(E(W|T)) = EW = \tau(\theta)$

"iterative" property of expected value

(hence $\hat{\tau}(T)$ is unbiased for $\tau(\theta)$).

ii) We first show that "always" $Var(Y|X) \le Var\,Y$ (i.e. the variance is never increased after conditioning)

Let $a(X) = E(Y|X)$. Then:

$$Var\,Y = E(Y - a(X) + a(X) - EY)^2 = E(Y-a(X))^2 + E(a(X)-EY)^2 + 2E[(Y-a(X))(a(X)-EY)]$$

Next: $E[(Y-a(X))(a(X)-EY)] = E_X[E(Y-a(X))(a(X)-EY)|X]$

$= E_X\{(a(X)-EY)E(Y-a(X)|X)\} =$

$= E_X\{(a(X)-EY)(\underbrace{E(Y|X)}_{=a(X)}-a(X))\} = 0$

Hence $Var\,Y = E(Y-a(X))^2 + E(a(X)-EY)^2 \ge$

$\ge E(a(X)-EY)^2 = E(a(X)-E(a(X)))^2 = Var(a(X))$

i.e. $Var(Y|X) \le Var\,Y$

3). I discussed sufficiency and completeness and its role in justifying the Lehmann-Scheffe theorem which gives us the recipe to obtain UMVUE by Rao-Blackwellising an unbiased estimator by conditioning on complete and sufficient statistic. This discussion is in the lecture notes.

NEXT, I solved a variety of illustrative examples.

—continued—

# Some of my white board writing in week 6



Completeness, Lehmann-Scheffe:

• $X = (X_1, X_2, \dots, X_n)$ i.i.d. $N(0, \theta)$. Show, $T = \overline{X}$ is not complete for $\theta$. It suffices to find a counterexample. Here it is: take $g(t) = t \neq 0$. We have
$E_\theta g(T) = E_\theta(\overline{X}) = 0 \; \forall \theta > 0$ but $g(t) \not\equiv 0$

• $X = (X_1, X_2, \dots, X_n)$ i.i.d. Bernoulli with parameter $\theta$. The statistic $T = \sum_{i=1}^{n} X_i$ is complete for $\theta \in (0,1)$:
We know $T \sim Bin(n, \theta) \longrightarrow P_\theta(T=t) = \binom{n}{t}\theta^t(1-\theta)^{n-t}$

Take $E_\theta g(T) = 0 \;\; \forall \theta \in (0,1) \Rightarrow \sum_{t=0}^{n} g(t)\binom{n}{t}\theta^t(1-\theta)^{n-t} = 0$

$\Rightarrow \underbrace{(1-\theta)^n}_{\neq 0} \cdot \sum_{t=0}^{n} g(t)\binom{n}{t}\eta^t = 0 \quad \forall \; \eta = \frac{\theta}{1-\theta} \in (0, \infty)$.

Then all coefficients $g(t)\binom{n}{t} = 0$ must hold and
Since $\binom{n}{t} \neq 0 \Rightarrow g(t) = 0, \; t = 0, 1, 2, \dots, n \Rightarrow P_\theta(g(T) = 0) = 1$
and $T$ is complete.

We also know: $T$ is sufficient. Hence if we start with an unbiased estimator $W$ of $\tau(\theta) = \theta(1-\theta)$ and calculate $E(W|T)$ in a second step, we will get the UMVUE of $\tau(\theta) = \theta(1-\theta)$.
Suggestions for $W$: $W = X_1(1 - X_2)$, (or $\tilde{W} = I_{(X_1=1)}(X) I_{(X_2=0)}(X)$).
We see: $E_\theta W = E_\theta X_1 - E_\theta(X_1 X_2) = \theta - E_\theta X_1 E_\theta X_2 = \theta - \theta^2 = \theta(1-\theta)$
(Similarly $E_\theta \tilde{W} = EI_{(X_1=1)}(X) \cdot EI_{(X_2=0)}(X) =$
$= P(X_1 = 1) P(X_2 = 0) = \theta(1-\theta)$.

Now we get:

$$E_\theta(W|T=t) = 1 * P_\theta(W=1|T=t) + 0 =$$

$$= \frac{P(W=1 \cap T=t)}{P(T=t)} = \frac{P(X_1=1 \cap X_2=0 \cap \sum_{i=3}^{n} X_i = t-1)}{P(\sum_{i=1}^{n} X_i = t)} =$$

made sure to have intersection of independent events

$$= \frac{P(X_1=1)\,P(X_2=0)\,P(Bin(n-2,\theta)=t-1)}{P(Bin(n,\theta)=t)} = \frac{\theta(1-\theta)\binom{n-2}{t-1}\theta^{t-1}(1-\theta)^{n-t-1}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} =$$

$$= \frac{\binom{n-2}{t-1}}{\binom{n}{t}} = \boxed{\bar{X}(1-\bar{X})\frac{n}{n-1}}$$ being the UMVUE of $\tau(\theta)=\theta(1-\theta)$

• For uniform in $[0,\theta)$ distribution, $Y = \frac{n+1}{n}X_{(n)}$ is the UMVUE of the parameter $\tau(\theta)=\theta$. Justification:
From the previous lectures we know that $Y$ is unbiased for $\theta$ and that $X_{(n)}$ is a sufficient statistic for $\theta$. Now we show that $X_{(n)}$ is also complete. Recall that
$$f_{X_{(n)}}(t,\theta) = \begin{cases} \frac{nt^{n-1}}{\theta^n}, & 0<t<\theta \\ 0 & \text{else} \end{cases}$$ (derived last lecture)

Take $E_\theta g(X_{(n)}) = 0 \ \forall \theta \in (0,\infty) \Rightarrow \frac{n}{\theta^n}\int_0^\theta g(t)t^{n-1}dt = 0$

$0 = \frac{d}{d\theta}\int_0^\theta g(t)t^{n-1}dt = g(\theta)\theta^{n-1}$. But since $\theta>0$, this implies $g(\theta)=0$ for all $\theta>0$. Equivalently: $P_\theta(g(X_{(n)})=0)=1$ and $X_{(n)}$ is complete.

Now: $Y = \frac{n+1}{n}X_{(n)}$ is unbiased for $\theta$ and is a function of complete and sufficient statistic. Conditioning $Y$ on $X_{(n)}$ does not change it $E(Y/X_{(n)}) = \frac{n+1}{n}X_{(n)}$

Hence, by Lehmann-Scheffe: $Y = \frac{n+1}{n}X_{(n)}$ is UMVUE of $\tau(\theta)=\theta$.

• the Poisson example continued: For $\tau(\theta) = e^{-\theta}$, UMVUE was advertised as being $\left(1-\frac{1}{n}\right)^{n\bar{X}}$ and below I justify this claim:

First we note that $T = \sum_{i=1}^{n} X_i \sim \text{Poisson}(n\theta)$ (this is a known property of the Poisson distribution. $T = \sum_{i=1}^{n} X_i$ is known to be sufficient for $\theta$ from our previous lectures. Now we will show that it is also complete:

Take $E_\theta g(T) = 0 \ \forall \theta > 0 \Rightarrow \sum_{t=0}^{\infty} g(t) \frac{e^{-n\theta}(n\theta)^t}{t!} = 0$

for all $\theta > 0$. This means $\underset{\neq 0}{e^{-n\theta}} \cdot \underbrace{\sum_{t=0}^{\infty} g(t) \frac{(n\theta)^t}{t!}}_{} = 0 \ \forall \theta > 0$

This polynomial of $\theta$ must be then $= 0 \ \forall \theta > 0 \rightarrow$ the coefficients

$\frac{g(t) n^t}{t!}$ must be all $= 0$ which implies $g(t) = 0, t = 0, 1, \ldots$

i.e. $\quad P_\theta(g(T) = 0) = 1$ and $T = \sum_{i=1}^{n} X_i$ is complete.

To find an unbiased starting estimator for $\tau(\theta) = e^{-\theta}$ we use the interpretation $e^{-\theta} = P_\theta(X_1 = 0)$ of $\tau(\theta)$.

Hence $W = I_{(X_1 = 0)}(\mathbf{X})$ would be unbiased for $\tau(\theta)$:

$E_\theta W = 1 * P_\theta(X_1 = 0) = e^{-\theta}$. If we now condition on the complete & sufficient $T = \sum_{i=1}^{n} X_i$, we will get the UMVUE:

$$E_\theta(W \mid T = t) = 1 * P(W = 1 \mid T = t) = \frac{P(W = 1 \cap T = t)}{P(T = t)} =$$

$$= \frac{P(X_1 = 0 \cap \sum_{i=1}^{n} X_i = t)}{P(\sum_{i=1}^{n} X_i = t)} = \frac{P(X_1 = 0 \cap \sum_{i=2}^{n} X_i = t)}{P(\sum_{i=1}^{n} X_i = t)} =$$

$$= \frac{e^{-\theta} \cdot e^{-(n-1)\theta} \, ((n-1)\theta)^t / \cancel{t!}}{e^{-n\theta} (n\theta)^t / \cancel{t!}} = \left(\frac{n-1}{n}\right)^t = \left(1 - \frac{1}{n}\right)^{n\bar{X}} \quad \text{qed}$$

- I also justified why for the uniform distribution in $[0,\theta]$, the maximal observation $X_{(n)}$ is the MLE, i.e. $\hat{\theta}_{MLE} = X_{(n)}$.

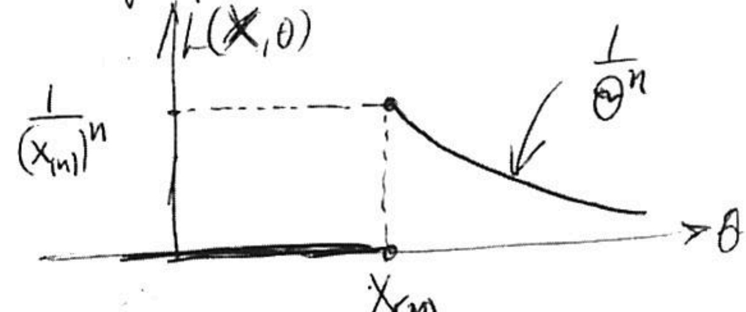
I noticed that $L(\underline{X},\theta) = \prod_{i=1}^{n} f(x_i,\theta)$ is not differentiable for all $\theta$, hence instead of trying to solve the equation $V(\underline{X}, \hat{\theta}_{MLE}) = 0$ to find the MLE (which is what we would do in "regular" cases), we look directly into the shape of $L(\underline{X},\theta)$ to see which is the argument that maximizes it.

Since $f(x,\theta) = \frac{1}{\theta} I_{(x,\infty)}(\theta)$ then

$$L(\underline{X},\theta) = \prod_{i=1}^{n} f(X_i,\theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I_{(X_i,\infty)}(\theta) = \boxed{\frac{1}{\theta^n} I_{(X_{(n)},\infty)}(\theta)}$$

using properties of indicators

If we now graph $L(\underline{X},\theta)$ we get after plugging the sample:




and clearly $L(\underline{X},\theta)$ is maximized when $\theta = X_{(n)}$, i.e. $\hat{\theta}_{MLE} = X_{(n)}$ by direct inspection

Note that the $\hat{\theta}_{MLE} = X_{(n)}$ is different from the UMVUE $Y = \frac{n+1}{n} X_{(n)}$ that we derived earlier.

Finally, I stressed that usually parameters of interest such as $\tau(\theta)=\theta(1-\theta)$ for the Bernoulli; $\tau(\theta)=e^{-\theta}=P(X_1=0)$ for the Poisson; $\tau(\theta)=\theta e^{-\theta}=P(X_1=1)$ for the Poisson, etc., typically have some probabilistic interpretation that can be exploited to suggest a (simple) unbiased estimator $W$ which then can be Rao-Blackwellized to obtain the UMVUE. The tutorial questions (set 2) contain a lot of such exercises. By doing them, you can get a feeling how to proceed in a particular situation.

ONE MORE REMARK I made at the end: If we have realized that we are dealing with a one-parameter exponential family density $f(x,\theta)=a(\theta)b(x)\exp(c(\theta)d(x))$ then the statistic $T(X)=\sum_{i=1}^{n} d(X_i)$ is complete and minimal sufficient. We do not need to separately check completeness for such families.