Some white board writing from week 5

I continued lecturing about inference principles. I discussed the weak likelihood principle. I also presented an example with 3 different experiments that give rise to proportional likelihoods. These discussions are presented in details in the notes and I abstain from reproducing them here.

Then I introduced the notion of (Fisher) Information as the variance of the score: Information in the sample $\mathbb{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$ of n i.i.d. observations from distribution with a density $f(x,\theta)$, $\theta \in R^1$ as

$$I_{\mathbb{X}}(\theta) = Var_\theta(V(\mathbb{X},\theta)) = E_\theta\left(\frac{\partial}{\partial\theta} \ln L(\mathbb{X},\theta)\right)^2 =$$

$$= E_\theta\left(\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f(X_i,\theta)\right)^2.$$

I also discussed the properties of the Information quantity introduced above such as:

i) additivity over independent samples

ii) preservation of information by a sufficient statistic, i.e. $I_T(\theta) = I_{\mathbb{X}}(\theta)$ when T is sufficient

iii) alternative way to calculate the information in the sample: under smoothness regularity conditions:

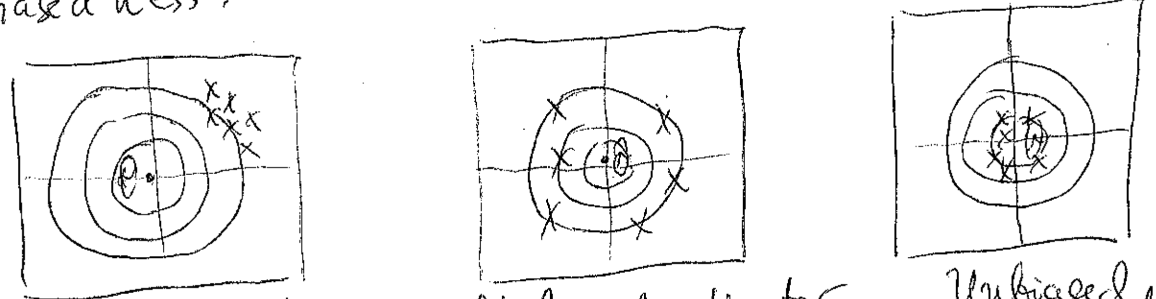
$$I_{\mathbb{X}}(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2} \ln L(\mathbb{X},\theta)\right)$$

iv) For <u>any</u> statistic $T(\mathbb{X})$ it holds $I_T(\theta) \le I_{\mathbb{X}}(\theta)$, with equality if and only if $T(\mathbb{X})$ is sufficient for θ.

I went through the proofs of all these statements but the proofs are thoroughly presented in the notes and I am not reproducing them here.

Then I started discussing unbiasedness and CR inequality.

1) First I discussed the relevance of the notion of unbiasedness:



Biased estimator — Unbiased estimator with high variance — Unbiased estimator with small variance

Thanks to the decomposition
$$MSE_\theta(T_n) = E_\theta(T_n - \theta \pm ET_n) = Var_\theta T_n + (b_n(\theta))^2$$
and the graphs above, it does make sense to look for the estimator with the smallest possible variance in the class of unbiased estimators.

I did make a cautious remark that sometimes an unbiased estimator may not be that useful. Take $f(x,\theta) = \theta(1-\theta)^{x-1}$, $x = 1, 2, \dots$ & $T(x)$ based on $n = 1$ observation would mean that $\sum_{x=1}^{\infty} T(x)\theta(1-\theta)^{x-1} = \theta$ holds $\forall \theta \in (0,1)$. Cancel θ, set $\eta = 1-\theta$ and we see:
$T(1) + \eta T(2) + \eta^2 T(3) + \dots = 1$ $\forall \eta \in (0,1)$ must hold. Hence $T(1) = 1$ and $T(2) = T(3) = T(4) = \dots = 0$!

Note: the estimator $\tilde{T}(x) = \frac{1}{x}$ (which happens to be the MLE in this example) is much more useful.

2) Example illustrating that when condition $(*)$ is violated, we could have estimators which are unbiased and have a variance $<$ CR bound:

Let $X_1, X_2, \dots, X_n$ be i.i.d. uniform in $[0,\theta)$ (so that the support of the density depends on θ & $(*)$ is violated). Then:
$$f_{X_1}(x) = \frac{1}{\theta} I_{(0,\theta)}(x) \longrightarrow F_{X_1}(x) = \begin{cases} 0 & x \le 0 \\ \frac{x}{\theta} & 0 < x \le \theta \\ 1 & x > \theta \end{cases}$$
$$F_{X_{(n)}}(y) = P(X_{(n)} \le y) = P(X_1 < y \cap X_2 < y \cap \dots X_n < y) = \left(\frac{y}{\theta}\right)^n, \quad 0 < y < \theta$$
Hence $f_{X_{(n)}}(y) = \frac{n y^{n-1}}{\theta^n}$, $0 < y < \theta$ (and zero else)

Take $EX_{(n)} = \int_0^\theta y \frac{ny^{n-1}}{\theta^n} dy = \frac{n}{n+1}\theta \neq \theta$,

i.e. $X_{(n)}$ is biased for estimating θ. BUT:

$T = \frac{n+1}{n} X_{(n)}$ is unbiased for estimating θ.

$$\operatorname{Var} T = E(T^2) - \theta^2 = \left(\frac{n+1}{n}\right)^2 \int_0^\theta y^2 \frac{ny^{n-1}}{\theta^n} dy - \theta^2 = \dots = \frac{\theta^2}{n(n+2)}$$

But for $f_{X_1}(\theta) = \frac{1}{\theta}, 0 < x < \theta$ we have $\ln f_{X_1}(\theta) = -\ln\theta$

$\frac{\partial}{\partial\theta} \ln f_{X_1}(\theta) = -\frac{1}{\theta}$ $\quad E\left(\frac{\partial}{\partial\theta}\ln f_{X_1}(\theta)\right)^2 = \frac{1}{\theta^2}$ and

reckless application of CR Bound would imply

CR Bound $= \frac{\theta^2}{n}$. As we see now:

$$\operatorname{Var}(T) = \frac{\theta^2}{n(n+2)} < \frac{\theta^2}{n}$$

Reason for this seeming contradition: the condition (*) was violated in this example!

3) Score function for the Poisson(θ) example:

$X_1, X_2, \dots, X_n$ i.i.d. Poisson(θ)

$$P(X_i = x) = \frac{e^{-\theta}\theta^x}{x!}, x = 0, 1, 2, \dots$$

$$L(\mathbf{X}, \theta) = \frac{e^{-n\theta}\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}. \quad \text{Hence}$$

$$V(\mathbf{X}, \theta) = \frac{\partial}{\partial\theta}\log L(\mathbf{X}, \theta) = -n + \frac{\sum_{i=1}^n X_i}{\theta}$$

If $\tau(\theta) = \theta \to V(\mathbf{X}, \theta) = \frac{n}{\theta}(\bar{X} - \theta) \to$ factorization possible and $\bar{X}$ is the UMVUE of θ that attains the CR Bound