Lecture 7

We showed that

$$\mathbb{E}_\theta(S - m(\theta))l'_n \leq (\mathbb{E}_\theta(S - m(\theta))^2)^{1/2}(nI(\theta))^{1/2}.$$

Next, let us compute the left hand side. We showed that $\mathbb{E}_\theta l'(X_1|\theta) = 0$ which implies that

$$\mathbb{E}_\theta l'_n = \sum \mathbb{E}_\theta l'(X_i|\theta) = 0$$

and, therefore,

$$\mathbb{E}_\theta(S - m(\theta))l'_n = \mathbb{E}_\theta S\ l'_n - m(\theta)\mathbb{E}_\theta l'_n = \mathbb{E}_\theta S\ l'_n.$$

Let $X = (X_1, \ldots, X_n)$ and denote by

$$\varphi(X|\theta) = f(X_1|\theta)\ldots f(X_n|\theta)$$

the joint p.d.f. (or likelihood) of the sample $X_1, \ldots, X_n$ We can rewrite l'_n in terms of this joint p.d.f. as

$$l'_n = \frac{\partial}{\partial\theta}\sum_{i=1}^{n}\log f(X_i|\theta) = \frac{\partial}{\partial\theta}\log\varphi(X|\theta) = \frac{\varphi'(X|\theta)}{\varphi(X|\theta)}.$$

Therefore, we can write

$$\begin{aligned}
\mathbb{E}_\theta S l'_n &= \mathbb{E}_\theta S(X)\frac{\varphi'(X|\theta)}{\varphi(X|\theta)} = \int S(X)\frac{\varphi'(X|\theta)}{\varphi(X|\theta)}\varphi(X)dX \\
&= \int S(X)\varphi'(X|\theta)dX = \frac{\partial}{\partial\theta}\int S(X)\varphi(X|\theta)dX = \frac{\partial}{\partial\theta}\mathbb{E}_\theta S(X) = m'(\theta).
\end{aligned}$$

Of course, we integrate with respect to all coordinates, i.e. $dX = dX_1\ldots dX_n$. We finally proved that

$$m'(\theta) \leq (\mathbb{E}_\theta(S - m(\theta))^2)^{1/2}(nI(\theta))^{1/2} = (\mathrm{Var}_\theta(S))^{1/2}(nI(\theta))^{1/2}$$

which implies Rao-Crámer inequality.

$$\mathrm{Var}_\theta(S) \geq \frac{(m'(\theta))^2}{nI(\theta)}.$$

The inequality will become equality only if there is equality in the Cauchy inequality applied to random variables

$$S - m(\theta) \text{ and } l'_n.$$

But this can happen only if there exists $t = t(\theta)$ such that

$$S - m(\theta) = t(\theta) l'_n = t(\theta) \sum_{i=1}^{n} l'(X_i|\theta).$$

7.1 Efficient estimators.

Definition: Consider statistic $S = S(X_1, \ldots, X_n)$ and let

$$m(\theta) = \mathbb{E}_\theta S(X_1, \ldots, X_n).$$

We say that S is an *efficient estimate* of $m(\theta)$ if

$$\mathbb{E}_\theta (S - m(\theta))^2 = \frac{(m'(\theta))^2}{nI(\theta)},$$

i.e. equality holds in Rao-Crámer's inequality.

In other words, efficient estimate S is the best possible unbiased estimate of $m(\theta)$ in a sense that it achieves the smallest possible value for the average squared deviation $\mathbb{E}_\theta (S - m(\theta))^2$ for all θ.

We also showed that equality can be achieved in Rao-Crámer's inequality only if

$$S = t(\theta) \sum_{i=1}^{n} l'(X_i|\theta) + m(\theta)$$

for some function $t(\theta)$. The statistic $S = S(X_1, \cdots, X_n)$ must a function of the sample only and it can not depend on θ. This means that efficient estimates do not always exist and they exist only if we can represent the derivative of log-likelihood l'_n as

$$l'_n = \sum_{i=1}^{n} l'(X_i|\theta) = \frac{S - m(\theta)}{t(\theta)},$$

where S does not depend on θ. In this case, S is an efficient estimate of $m(\theta)$.

Exponential-type families of distributions. Let us consider the special case of so called *exponential-type* families of distributions that have p.d.f. or p.f. $f(x|\theta)$ that can be represented as:

$$f(x|\theta) = a(\theta)b(x)e^{c(\theta)d(x)}.$$

In this case we have,

$$\begin{aligned} l'(x|\theta) &= \frac{\partial}{\partial\theta}\log f(x|\theta) = \frac{\partial}{\partial\theta}(\log a(\theta) + \log b(x) + c(\theta)d(x)) \\ &= \frac{a'(\theta)}{a(\theta)} + c'(\theta)d(x). \end{aligned}$$

This implies that

$$\sum_{i=1}^n l'(X_i|\theta) = n\frac{a'(\theta)}{a(\theta)} + c'(\theta)\sum_{i=1}^n d(X_i)$$

and

$$\frac{1}{n}\sum_{i=1}^n d(X_i) = \frac{1}{nc'(\theta)}\sum_{i=1}^n l'(X_i|\theta) - \frac{a'(\theta)}{a(\theta)c'(\theta)}.$$

If we take

$$S = \frac{1}{n}\sum_{i=1}^n d(X_i) \text{ and } m(\theta) = \mathbb{E}_\theta S = -\frac{a'(\theta)}{a(\theta)c'(\theta)}$$

then S will be an efficient estimate of $m(\theta)$.

Example. Consider a family of Poisson distributions $\Pi(\lambda)$ with p.f.

$$f(x|\lambda) = \frac{\lambda^x}{x!}e^{-\lambda} \text{ for } x = 0, 1, \ldots$$

This can be expressed as exponential-type distribution if we write

$$\frac{\lambda^x}{x!}e^{-\lambda} = \underbrace{e^{-\lambda}}_{a(\lambda)}\underbrace{\frac{1}{x!}}_{b(x)}\exp\Big\{\underbrace{\log\lambda}_{c(\lambda)}\underbrace{x}_{d(x)}\Big\}.$$

As a result,

$$S = \frac{1}{n}\sum_{i=1}^n d(X_i) = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}$$

is efficient estimate of its expectation $m(\lambda) = \mathbb{E}_\lambda S = \mathbb{E}_\lambda X_1 = \lambda$. We can also compute its expectation directly using the formula above:

$$\mathbb{E}_\lambda S = -\frac{a'(\lambda)}{a(\lambda)c'(\lambda)} = \frac{-(-e^{-\lambda})}{e^{-\lambda}(\frac{1}{\lambda})} = \lambda.$$

Maximum likelihood estimators. Another interesting consequence of Rao-Crámer's theorem is the following. Suppose that the MLE $\hat{\theta}$ is unbiased:

$$\mathbb{E}\hat{\theta} = \theta.$$

If we take $S = \hat{\theta}$ and $m(\theta) = \theta$ then Rao-Crámer's inequality implies that

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{nI(\theta)}.$$

On the other hand when we showed asymptotic normality of the MLE we proved the following convergence in distribution:

$$\sqrt{n}(\hat{\theta} - \theta) \to N\Big(0, \frac{1}{I(\theta)}\Big).$$

In particular, the variance of $\sqrt{n}(\hat{\theta} - \theta)$ converges to the variance of the normal distribution $1/I(\theta)$, i.e.

$$\mathrm{Var}(\sqrt{n}(\hat{\theta} - \theta)) = n\mathrm{Var}(\hat{\theta}) \to \frac{1}{I(\theta)}$$

which means that Rao-Crámer's inequality becomes equality in the limit. This property is called the *asymptotic efficiency* and we showed that unbiased MLE is asymptotically efficient. In other words, for large sample size n it is almost best possible.