Lecture 6

Let us compute Fisher information for some particular distributions.

Example 1. The family of Bernoulli distributions $B(p)$ has p.f.

$$f(x|p) = p^x(1-p)^{1-x}$$

and taking the logarithm

$$\log f(x|p) = x\log p + (1-x)\log(1-p).$$

The second derivative with respect to parameter p is

$$\frac{\partial}{\partial p}\log f(x|p) = \frac{x}{p} - \frac{1-x}{1-p},\ \ \frac{\partial^2}{\partial p^2}\log f(x|p) = -\frac{x}{p^2} - \frac{1-x}{(1-p)^2}$$

then we showed that Fisher information can be computed as:

$$I(p) = -\mathbb{E}\frac{\partial^2}{\partial p^2}\log f(X|p) = \frac{\mathbb{E}X}{p^2} + \frac{1-\mathbb{E}X}{(1-p)^2} = \frac{p}{p^2} + \frac{1-p}{(1-p)^2} = \frac{1}{p(1-p)}.$$

The MLE of p is $\hat{p} = \bar{X}$ and the asymptotic normality result from last lecture becomes

$$\sqrt{n}(\hat{p} - p_0) \to N(0, p_0(1-p_0))$$

which, of course, also follows directly from the CLT.

Example. The family of exponential distributions $E(\alpha)$ has p.d.f.

$$f(x|\alpha) = \begin{cases} \alpha e^{-\alpha x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

and, therefore,

$$\log f(x|\alpha) = \log\alpha - \alpha x \Rightarrow \frac{\partial^2}{\partial\alpha^2}\log f(x|\alpha) = -\frac{1}{\alpha^2}.$$

This does not depend on X and we get

$$I(\alpha) = -\mathbb{E}\frac{\partial^2}{\partial\alpha^2}\log f(X|\alpha) = \frac{1}{\alpha^2}.$$

Therefore, the MLE $\hat{\alpha} = 1/\bar{X}$ is asymptotically normal and

$$\sqrt{n}(\hat{\alpha} - \alpha_0) \to N(0, \alpha_0^2).$$

□

6.1 Rao-Crámer inequality.

Let us start by recalling the following simple result from probability (or calculus).

Lemma. (Cauchy inequality) *For any two random variables X and Y we have:*

$$\mathbb{E}XY \leq (\mathbb{E}X^2)^{1/2}(\mathbb{E}Y^2)^{1/2}.$$

The inequality becomes equality if and only if $X = tY$ for some $t \geq 0$ with probability one.

Proof. Let us consider the following function

$$\varphi(t) = \mathbb{E}(X - tY)^2 = \mathbb{E}X^2 - 2t\mathbb{E}XY + t^2\mathbb{E}Y^2 \geq 0.$$

Since this is a quadractic function of t, the fact that it is nonnegative means that it has not more than one solution which is possible only if the discriminant is non positive:

$$D = 4(\mathbb{E}XY)^2 - 4\mathbb{E}Y^2\mathbb{E}X^2 \leq 0$$

and this implies that

$$\mathbb{E}XY \leq (\mathbb{E}X^2)^{1/2}(\mathbb{E}Y^2)^{1/2}.$$

Also $\varphi(t) = 0$ for some t if and only if $D = 0$. On the other hand, $\varphi(t) = 0$ means

$$\mathbb{E}(X - tY)^2 = 0 \Rightarrow X = tY$$

with probability one.

□

Let us consider statistic

$$S = S(X_1, \ldots, X_n)$$

which is a function of the sample $X_1, \ldots, X_n$. Let us define a function

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

where $\mathbb{E}_\theta$ is the expectation with respect to distribution $\mathbb{P}_\theta$. In other words, $m(\theta)$ denotes the mean of S when the sample has distribution $\mathbb{P}_\theta$. The following is the main result of this lecture.

Theorem. (The Rao-Crámer inequality). *We have,*

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

This inequality becomes equality if and only if

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

for some function $t(\theta)$ *and where* $l(X|\theta) = \log f(X|\theta)$.

Proof: Let us introduce the notation

$$l(x|\theta) = \log f(x|\theta)$$

and consider a function

$$l_n = l_n(X_1, \ldots, X_n, \theta) = \sum_{i=1}^{n} l(X_i|\theta).$$

Let us apply Cauchy inequality in the above Lemma to the random variables

$$S - m(\theta) \text{ and } l_n' = \frac{\partial l_n}{\partial \theta}.$$

We have:

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

Let us first compute $\mathbb{E}_\theta(l_n')^2$. If we square out $(l_n')^2$ we get

$$\begin{aligned}
\mathbb{E}_\theta(l_n')^2 &= \mathbb{E}_\theta(\sum_{i=1}^{n} l'(X_i|\theta))^2 = \mathbb{E}_\theta \sum_{i=1}^{n}\sum_{j=1}^{n} l'(X_i|\theta)l'(X_j|\theta) \\
&= n\mathbb{E}_\theta(l'(X_1|\theta))^2 + n(n-1)\mathbb{E}_\theta l(X_1|\theta)\mathbb{E}_\theta l(X_2|\theta)
\end{aligned}$$

where we simply grouped n terms for $i = j$ and remaining $n(n-1)$ terms for $i \neq j$. By definition of Fisher information

$$I(\theta) = \mathbb{E}_\theta(l'(X_1|\theta))^2.$$

Also,

$$\begin{aligned}
\mathbb{E}_\theta l'(X_1|\theta) &= \mathbb{E}_\theta \frac{\partial}{\partial\theta} \log f(X_1|\theta) = \mathbb{E}_\theta \frac{f'(X_1|\theta)}{f(X_1|\theta)} = \int \frac{f'(x|\theta)}{f(x|\theta)} f(x|\theta) dx \\
&= \int f'(x|\theta) dx = \frac{\partial}{\partial\theta} \int f(x|\theta) dx = \frac{\partial}{\partial\theta} 1 = 0.
\end{aligned}$$

We used here that $f(x|\theta)$ is a p.d.f. and it integrates to one. Combining these two facts, we get

$$\mathbb{E}_\theta (l_n')^2 = nI(\theta).$$