LECTURE 11

EXPONENTIAL FAMILY, FISHER INFORMATION, CRAMER-RAO LOWER BOUND (CRLB)

EXPONENTIAL FAMILY PDFS

- *X* is a continuous (discrete) rv with pdf $f(x;\theta)$, $\theta \in \Omega$. If the pdf can be written in the following form

$$f(x;\theta) = h(x)c(\theta)\exp\left(\sum_{j=1}^{k} w_j(\theta)t_j(x)\right)$$

then, the pdf is a member of exponential class of pdfs of the continuous (discrete) type. (Here, k is the number of parameters)

REGULAR CASE OF THE EXPONENTIAL FAMILY

- We have a regular case of the exponential class of pdfs of the continuous type if

a) Range of X does not depend on θ.

b) $c(\theta) \geq 0$, $w_1(\theta),\ldots, w_k(\theta)$ are real valued functions of θ for $\theta \in \Omega$.

c) $h(x) \geq 0$, $t_1(x),\ldots, t_k(x)$ are real valued functions of x.

If the range of X depends on θ, then it is called *irregular* exponential class or *range-dependent* exponential class.

EXAMPLE

X~Bin(n,p), where n is known. Is this pdf a member of exponential class of pdfs?

$$f(x;p)=\binom{n}{x}p^x(1-p)^{n-x}; \quad x=0,1,\ldots,n; \quad 0<p<1$$

$$=\binom{n}{x}(1-p)^n\exp(x\ln(\frac{p}{1-p}))$$

$$h(x)=\binom{n}{x} \quad \mathit{for} \quad x=0,\ldots,n; \quad c(p)=(1-p)^n \quad \mathit{for} \quad 0<p<1$$

$$t_1(x)=x \quad \mathit{for} \quad x=0,\ldots,n; \quad w_1(p)=\ln(\frac{p}{1-p}) \quad \mathit{for} \quad 0<p<1$$

Binomial family is a member of exponential family of distributions.

EXAMPLE

X~Cauchy(1,θ). Is this pdf a member of exponential class of pdfs?

$$f(x;\theta) = (\pi(1+[x-\theta]^2))^{-1} = \frac{1}{\pi}\exp\{-\ln(1+x^2-2\theta x+\theta^2)\}$$

$$h(x) = \frac{1}{\pi}; \quad c(\theta) = 1; \quad -\ln(1+x^2-2\theta x+\theta^2) \neq t_1(x)w_1(\theta)$$

Cauchy is not a member of exponential family.

EXPONENTIAL CLASS and CSS

- Random Sample from Regular Exponential Class

↓

$$Y = \sum_{i=1}^{n} t_j(X_i) \quad \text{is a css for } \theta.$$

↓

If Y is an UE of θ, Y is the UMVUE of θ.

EXAMPLE

Let X1,X2,…~Bin(1,p), i.e., Ber(p).

This family is a member of exponential family of distributions.

$$t_1(x) = x \quad for \quad x = 0,\ldots,n$$ is a CSS for p.

$$\sum_{i=1}^{n} t_1(x_i) = \sum_{i=1}^{n} x_i$$

$\bar{X}$ is UE of p and a function of CSS.

$\Rightarrow$ $\bar{X}$ is UMVUE of p.

EXAMPLES

$X \sim N(\mu, \sigma^2)$ where both μ and σ^2 is unknown. Find a css for μ and σ^2.

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2}} e^{-\frac{\mu^2}{2\sigma^2}} e^{\left(-\frac{x^2}{2\sigma^2}+\frac{\mu x}{\sigma^2}\right)}$$

$$\theta_1 = \frac{\mu^2}{2\sigma^2}, \quad \theta_2 = \frac{1}{\sigma^2}, \quad h(x) = \frac{1}{\sqrt{2\pi}}, \quad c(\theta) = e^{-\theta_1}\sqrt{\theta_2}$$

$$w_1 = 2\theta_1, \quad w_2 = -\frac{1}{2}\theta_2, \quad t_1 = x, \quad t_2 = x$$

$$t_1(x) = x \quad for \quad x = 0,\ldots,n$$

$$\sum_{i=1}^{n} t_1(x_i) = \sum_{i=1}^{n} x_i$$

$$t_2(x) = x^2 \quad for \quad x = 0,\ldots,n$$

$$\sum_{i=1}^{n} t_2(x_i) = \sum_{i=1}^{n} x^2{}_i$$

are css for μ and σ^2

THE SCORE

- The score of the family $f(x|\theta)$ is the random variable

$$\frac{\partial}{\partial\theta}\ln f(x|\theta) = \frac{\frac{\partial}{\partial\theta} f(x|\theta)}{f(x|\theta)}$$

measures the "sensitivity" of $f(x|\theta)$ as a function of the parameter θ .

$$E[\frac{\partial}{\partial\theta}\ln f(x|\theta)]=0$$

Proof

$$E[\frac{\partial}{\partial\theta}\ln f(x|\theta)]=\int\frac{\frac{\partial}{\partial\theta}f(x|\theta)}{f(x|\theta)}f(x|\theta)dx=\int\frac{\partial}{\partial\theta}f(x|\theta)dx$$

$$=\frac{\partial}{\partial\theta}\int f(x|\theta)dx=\frac{\partial}{\partial\theta}1=0$$

As a result

$$\operatorname{var}\left[\frac{\partial}{\partial\theta}\ln f(x|\theta\right]=E\left[\left(\frac{\partial}{\partial\theta}\ln f(x|\theta)-E\left[\frac{\partial}{\partial\theta}\ln f(x|\theta)\right]\right)^2\right]=E\left[\left(\frac{\partial}{\partial\theta}\ln f(x|\theta)\right)^2\right]$$

Example

- Consider the normal distribution $N(\mu,1)$

$$f(x\,|\,\mu)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(x-\mu)^2\right)$$

$$\ln f(x\,|\,\mu)=-\frac{1}{2}\ln(2\pi)-\frac{1}{2}(x-\mu)^2$$

$$\text{The score}=s=\frac{\partial}{\partial\mu}\ln f(x\,|\,\mu)=x-\mu$$

- clearly, $E[s]=E[x-\mu]=E[x]-\mu=0$
- and $\text{var}(s)=E[s^2]=E[(x-\mu)^2]=\sigma^2=1$

THE SCORE - VECTOR FORM

- In case where $\theta = (\theta_1, \ldots, \theta_k)$ is a vector, the score S is the vector whose ith component is

$$s_i = \frac{\partial}{\partial \theta_i} \ln f(x \mid \theta)$$

Example

$$f(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2 \right)$$

$$\ln f(x \mid \mu, \sigma) = -\frac{1}{2}\ln(2\pi) - \ln \sigma - \frac{1}{2\sigma^2}(x-\mu)^2$$

$$\frac{\partial}{\partial \mu} \ln f(x \mid \mu, \sigma) = \frac{x-\mu}{\sigma^2}$$

$$\frac{\partial}{\partial \sigma} \ln f(x \mid \mu, \sigma) = -\frac{1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3}$$

$$S = \left(\frac{x-\mu}{\sigma^2}, -\frac{1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3} \right)$$

FISHER INFORMATION

- Fisher information (about θ), is the variance of the score

$$J(\theta) = E\left[\frac{\partial}{\partial\theta}\ln p(x|\theta)\right]^2$$

- It is designed to provide a measure of how much information the parametric probability law $f(x|\theta)$ carries about the θ
- The properties:
  - The larger the sensitivity of $f(x|\theta)$ to changes in θ, the larger should be the information
  - The information carried by the combined law $f(x_1, x_2|\theta)$ should be the sum of those carried by $f(x_1|\theta)$ and $f(x_2|\theta)$
  - The information should be insensitive to the sign of the change in θ and preferably positive.
  - The information should be a deterministic quantity

Example

- Consider a random variable $X \sim N(\theta, \sigma^2)$

$$\ln f(x \mid \theta, \sigma) = -\frac{1}{2}\ln(2\pi) - \ln\sigma - \frac{1}{2\sigma^2}(x-\theta)^2$$

$$s = \frac{\partial}{\partial\theta}\ln p(x \mid \theta, \sigma) = \frac{x-\theta}{\sigma^2}$$

$$J(\theta) = E\left[s^2\right] = E\left[\left(\frac{x-\theta}{\sigma^2}\right)^2\right] = \frac{1}{\sigma^4}E\left[(x-\theta)^2\right] = \frac{\sigma^2}{\sigma^4} = 1/\sigma^2$$

- Whenever $\theta = (\theta_1, \ldots, \theta_k)$ is a vector, Fisher information is the <u>matrix</u> $J(\theta) = \left(J_{i,j}(\theta)\right)$ where

$$J_{i,j}(\theta) = \mathrm{cov}_\theta\left(\frac{\partial}{\partial\theta_i}\log f(x\,|\,\theta),\ \frac{\partial}{\partial\theta_j}\log f(x\,|\,\theta)\right)$$

- Let $x^{(n)} = x_1, \ldots, x_n$ be i.i.d. random variables $x_i \sim f(x_i\,|\,\theta)$. The score of $f(x^{(n)}\,|\,\theta)$ is the sum of the individual scores.

$$\begin{aligned} s(x^{(n)}) &= \frac{\partial}{\partial\theta}\ln f(x^{(n)}\,|\,\theta) = \frac{\partial}{\partial\theta}\ln\prod_i f(x_i\,|\,\theta) \\ &= \sum_i \frac{\partial}{\partial\theta}\ln f(x_i\,|\,\theta) \\ &= \sum_i s(x_i) \end{aligned}$$

- Based on n i.i.d. samples, the Fisher information about θ is

$$J_n(\theta) = E\left[\frac{\partial}{\partial\theta}\ln f(x^{(n)} \mid \theta)\right]^2$$

$$= E\left[s^2(x^{(n)})\right] = E\left[\sum_{i=1}^{n} s(x_i)\right]^2$$

$$= \sum_{i=1}^{n} E\left[s^2(x_i)\right] = nJ(\theta)$$

- Thus, the Fisher information is <u>additive</u> w.r.t. i.i.d. random variables.

Example

- If $x^{(n)} = x_1, \ldots, x_n$ are i.i.d. $x_i \sim N(\theta, \sigma^2)$, the score is

$$n \frac{\partial}{\partial \theta} \ln f(x \mid \theta, \sigma) = n \frac{x - \theta}{\sigma^2}$$

$$J(\theta) = 1/\sigma^2$$

$$J_n(\theta) = n/\sigma^2$$

CRAMER-RAO LOWER BOUND (CRLB)

- Theorem: Let $\hat{\theta}$ be an unbiased estimator for θ. Then

$$\operatorname{var}(\hat{\theta}) \geq \frac{1}{J(\theta)}$$

- Proof: Using $E(s)=0$ we have:

$$E\left[\left(s-E(s)\right)\left(\hat{\theta}-E(\hat{\theta})\right)\right] = E\left[s\left(\hat{\theta}-E(\hat{\theta})\right)\right]$$

$$= E\left[s\hat{\theta}\right] - E(\hat{\theta})E(s)$$

$$= E[s\hat{\theta}]$$

- Now

$$E\left[s\hat{\theta}\right] = \int \frac{\frac{\partial}{\partial\theta} f(x|\theta)}{f(x|\theta)} \hat{\theta} f(x|\theta) dx$$

$$= \int \frac{\partial}{\partial\theta} f(x|\theta) \hat{\theta} dx$$

$$= \frac{\partial}{\partial\theta} \int f(x|\theta) \hat{\theta} dx$$

$$= \frac{\partial}{\partial\theta} E_\theta\left[\hat{\theta}\right] = \frac{\partial}{\partial\theta} \theta = 1 \quad \text{If } \hat{\theta} \text{ is unbiased estimator}$$

- So, $E\left[\left(s-E(s)\right)\left(\hat{\theta}-E(\hat{\theta})\right)\right]=E[s\hat{\theta}]=1$
- By the Cauchy-Schwarz inequality

$$1=\left(E\left[\left(s-E(s)\right)\left(\hat{\theta}-E(\hat{\theta})\right)\right]\right)^2 \le E\left[\left(s-E(s)\right)^2\right]E\left[\left(\hat{\theta}-E(\hat{\theta})\right)^2\right]$$

$$=E\left[s^2\right]\mathrm{var}(\hat{\theta})$$

$$=J(\theta)\,\mathrm{var}(\hat{\theta})$$

- Therefore,

$$\mathrm{var}(\hat{\theta}) \ge \frac{1}{J(\theta)}$$

- For a biased estimator we have:

$$\mathrm{var}(\hat{\theta}) \ge \frac{\left(1+\frac{\partial}{\partial\theta}(E(\hat{\theta})-\theta)\right)^2}{J(\theta)}$$

CRAMER-RAO LOWER BOUND (CRLB) GENERAL

- Let $X_1, X_2, \ldots, X_n$ be sample random variables.
- The Fisher Information in the random sample is $J_n(\theta)$
- Range of X does not depend on θ.
- $Y=U(X_1, X_2, \ldots, X_n)$: a statistic; doesnot contain θ.
- Let $E(Y)=m(\theta)$.

$$V\left(Y\right) \geq \frac{\left[m'\left(\theta\right)\right]^2}{J_n(\theta)} \Rightarrow \text{The Cramer-Rao Lower Bound}$$

Example

- Let $x^{(n)} = x_1, \ldots, x_n$ be i.i.d. $x_i \sim N(\theta, \sigma^2)$ .From previous example $J_n(\theta) = n / \sigma^2$
- Now let $\hat{\theta}(x^{(n)}) = \frac{1}{n}\sum_{i=1}^{n} x_i$ be an (unbiased) estimator for θ .

$$E(\hat{\theta}^2) = \frac{1}{n^2} E\left(\sum_{i=1}^{n} x_i\right)^2 = \frac{1}{n^2}\left(n^2\theta^2 + n\sigma^2\right)$$

$$= \theta^2 + \sigma^2 / n$$

$$\mathrm{var}(\hat{\theta}) = E_\theta\left(\hat{\theta} - E(\hat{\theta})\right)^2 = E\left(\hat{\theta} - \theta\right)^2 = E(\hat{\theta}^2) - 2\theta E(\hat{\theta}) + \theta^2 = E(\hat{\theta}^2) - \theta^2$$

- So $\mathrm{var}(\hat{\theta}) = \sigma^2 / n$ matches the Cramer-Rao lower bound.

Example

- Suppose $x \sim \text{Binomial}(n, p)$

$$f(x;p) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$\ln f(x;p) = \ln \binom{n}{x} + x \ln p + (n-x)\ln(1-p)$$

$$\frac{\partial \ln f(x;p)}{\partial p} = \frac{x}{p} - \frac{n-x}{1-p} = \frac{x-np}{p(1-p)}$$

$$\left(\frac{\partial \ln f(x;p)}{\partial p}\right)^2 = \left(\frac{x-np}{p(1-p)}\right)^2$$

$$E\left[\left(\frac{\partial \ln f(x;p)}{\partial p}\right)^2\right] = \frac{E[(x-np)^2]}{p^2(1-p)^2} = \frac{\text{var}(X)}{p^2(1-p)^2} = \frac{np(1-p)}{p^2(1-p)^2} = \frac{n}{p(1-p)}$$

Any unbiased estimator $\hat{p}$ of p is efficient if satisfies

$$\operatorname{var}(\hat{p}) = \frac{1}{\frac{n}{p(1-p)}} = \frac{p(1-p)}{n} \; CRLB$$

Suppose $\hat{p} = \frac{x}{n}$

$$E(\hat{p}) = E\left(\frac{x}{n}\right) = \frac{E(x)}{n} = \frac{np}{n} = p \quad \text{unbiased}$$

$$\operatorname{var}(\hat{p}) = \operatorname{var}\left(\frac{x}{n}\right) = \frac{\operatorname{var}(x)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

Then $\hat{p}$ EE

LIMITING DISTRIBUTION OF MLEs

- $\hat{\theta}$: MLE of θ
- $X_1, X_2, \ldots, X_n$ is a random sample.

$$m(\hat{\theta}) \overset{assmytotically}{\sim} N(m(\theta), CRLB = \frac{\left[m'(\theta)\right]^2}{J_n(\theta)})$$

$$\text{for large n} \quad m(\hat{\theta}) \overset{assmytotically}{\sim} N(m(\theta), CRLB = \frac{1}{J_n(\theta)})$$

EFFICIENT ESTIMATOR

- $\hat{\theta}$ is an *efficient* estimator (EE) of θ if
  - $\hat{\theta}$ is UE of θ, and,
  - Var($\hat{\theta}$)=CRLB
- *Y* is an *efficient* estimator (EE) of its expectation, *m(θ)*, if its variance reaches the CRLB.
- An EE of *m(θ)* may not exist.
- The EE of *m(θ),* if exists, is unique.
- The EE of *m(θ)* is the unique MVUE of m(θ).

ASYMPTOTIC EFFICIENT ESTIMATOR

- *Y* is an asymptotic EE of $m(\theta)$ *if*

$$\lim_{n\to\infty} E\left(Y\right) = m\left(\theta\right)$$

and

$$\lim_{n\to\infty} V\left(Y\right) = CRLB$$