Statistics
math 414

Lecture 13

Vivek Tewary

Lemma. Define $I(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right]$

Under smoothness conditions on f, $I(\theta)$ can be written as

$$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right].$$

I is called the <u>Fisher information</u>.

Proof

$$\int f(x;\theta) = 1$$

$$\Rightarrow \quad \frac{\partial}{\partial\theta} \int f(x;\theta) = 0$$

But $\dfrac{\partial}{\partial\theta} f(x;\theta) = \left(\dfrac{\partial}{\partial\theta} \log f(x;\theta)\right) f(x;\theta)$

Proof

$$\int f(x;\theta) = 1$$

$$\Rightarrow \quad \frac{\partial}{\partial \theta} \int f(x;\theta) = 0$$

But $\displaystyle \frac{\partial}{\partial \theta} f(x;\theta) = \left(\frac{\partial}{\partial \theta} \log f(x;\theta) \right) f(x;\theta)$

Therefore, $\displaystyle \int \left(\frac{\partial}{\partial \theta} \log f(x;\theta) \right) f(x;\theta) = 0$

Differentiate w.r.t. θ again to get:

Proof

$$\int f(x;\theta) = 1$$

$$\Rightarrow \quad \frac{\partial}{\partial\theta}\int f(x;\theta) = 0$$

But $\frac{\partial}{\partial\theta} f(x;\theta) = \left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right) f(x;\theta)$

Therefore, $\int\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right) f(x;\theta) = 0$

Differentiate w.r.t. θ again to get:

$$\int\left(\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right) f(x;\theta) + \int\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2 f(x;\theta) = 0.$$

Proof

$$\int f(x;\theta) = 1$$

$$\Rightarrow \frac{\partial}{\partial\theta}\int f(x;\theta) = 0$$

But $\frac{\partial}{\partial\theta} f(x;\theta) = \left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right) f(x;\theta)$

Therefore, $\int\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right) f(x;\theta) = 0$

$E\left[\left(\frac{\partial}{\partial\theta}\log f\right)^2\right]$

Differentiate w.r.t. θ again to get:

$$\int\left(\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right) f(x;\theta) + \int\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2 f(x;\theta) = 0.$$

Proof

$$\int f(x;\theta) = 1$$

$$\Rightarrow \frac{\partial}{\partial\theta}\int f(x;\theta) = 0$$

But $\frac{\partial}{\partial\theta} f(x;\theta) = \left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right) f(x;\theta)$

Therefore, $\int\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right) f(x;\theta) = 0$

Differentiate w.r.t. θ again to get.

$$E\left[\frac{\partial^2}{\partial\theta^2}\log f\right]$$

$$\int\left(\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right) f(x;\theta) + \int\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2 f(x;\theta) = 0.$$

$$\int f(x;\theta) = 1$$

$$\Rightarrow \quad \frac{\partial}{\partial\theta} \int f(x;\theta) = 0$$

But $\quad \dfrac{\partial}{\partial\theta} f(x;\theta) = \left(\dfrac{\partial}{\partial\theta} \log f(x;\theta)\right) f(x;\theta)$

Notice that these two imply that

$$E\left[\frac{\partial}{\partial\theta} \log f(x;\theta)\right] = 0$$

$$\int f(x;\theta) = 1$$

$$\Rightarrow \quad \frac{\partial}{\partial\theta} \int f(x;\theta) = 0$$

But $\frac{\partial}{\partial\theta} f(x;\theta) = \left(\frac{\partial}{\partial\theta} \log f(x;\theta)\right) f(x;\theta)$

Notice that these two imply that

$$E\left[\frac{\partial}{\partial\theta} \log f(x;\theta)\right] = 0$$

Therefore $\boxed{I(\theta) = Var\left(\frac{\partial}{\partial\theta} \log f(x;\theta)\right)}$

Interpretation

The function $I(\theta)$ is the weighted mean of $-\dfrac{\partial^2 \log f(x,\theta)}{\partial \theta^2}$ with respect to the weight $f(x;\theta)$. If this is large then we have more information on θ.

Example Information for Bernoulli (θ)

$$\log f(x;\theta) = x \log \theta + (1-x) \log (1-\theta)$$

Example Information for Bernoulli (θ)

$$\log f(x;\theta) = x \log \theta + (1-x) \log (1-\theta)$$

$$\frac{\partial \log f(x;\theta)}{\partial \theta} = \frac{x}{\theta} - \frac{(1-x)}{(1-\theta)}$$

$$\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2} = \frac{-x}{\theta^2} - \frac{(1-x)}{(1-\theta)^2}$$

Example Information for Bernoulli (θ)

$$\log f(x;\theta) = x \log \theta + (1-x) \log(1-\theta)$$

$$\frac{\partial \log f(x;\theta)}{\partial \theta} = \frac{x}{\theta} - \frac{(1-x)}{(1-\theta)}$$

$$\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2} = \frac{-x}{\theta^2} - \frac{(1-x)}{(1-\theta)^2}$$

$$\therefore\ I(\theta) = -E\left[\frac{-X}{\theta^2} - \frac{(1-X)}{(1-\theta)^2}\right]$$

Example Information for Bernoulli (θ)

$$\log f(x;\theta) = x\log\theta + (1-x)\log(1-\theta)$$

$$\frac{\partial \log f(x;\theta)}{\partial\theta} = \frac{x}{\theta} - \frac{(1-x)}{(1-\theta)}$$

$$\frac{\partial^2 \log f(x;\theta)}{\partial\theta^2} = \frac{-x}{\theta^2} - \frac{(1-x)}{(1-\theta)^2}$$

$$\therefore\ I(\theta) = -E\left[\frac{-X}{\theta^2} - \frac{(1-X)}{(1-\theta)^2}\right]$$

$$= \frac{E[X]}{\theta^2} + \frac{E[1-X]}{(1-\theta)^2} = \frac{\theta}{\theta^2} + \frac{(1-\theta)}{(1-\theta)^2} = \frac{1}{\theta(1-\theta)}$$

What is the Fisher information of a sample of size n?

What is the Fisher information of a sample of size n?

Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from a distribution $f(x;\theta)$. The likelihood $L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$

Therefore, $\dfrac{\partial \log L(\theta;X)}{\partial\theta} = \sum_{i=1}^{n} \dfrac{\partial \log f(X_i;\theta)}{\partial\theta}$

$= \boxed{n\,I(\theta)}$

Theorem (Rao-Cramér Lower Bound) Let $X_1, X_2, \ldots, X_n$ be iid with common pdf $f(x;\theta)$ for $\theta \in \Omega$. Assume certain unstate conditions.

Let $Y = u(X_1, X_2, \ldots, X_n)$ be a statistic with mean $E(Y) = E[u(X_1, X_2, \ldots, X_n)] = k(\theta)$.

Then

$$\operatorname{Var}(Y) \geq \frac{(k'(\theta))^2}{n I(\theta)}.$$

Proof Assume that X is a continuous r.v.

$$E(Y) = k(\theta) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_1, x_2, \ldots x_n) \prod_{i=1}^{n} f(x_i; \theta)\, dx_1\, dx_2 \ldots dx_n$$

Proof Assume that X is a continuous r.v.

$$E(Y) = k(\theta) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_1, x_2, \ldots x_n) \prod_{i=1}^{n} f(x_i;\theta)\, dx_1\, dx_2 \ldots dx_n$$

Differentiating w.r.t. θ (this requires strong conditions on u, f)

$$k'(\theta) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_1, x_2, \ldots, x_n) \left(\sum_{i=1}^{n} \frac{1}{f(x_i;\theta)} \frac{\partial f(x_i;\theta)}{\partial \theta} \right) \prod_{i=1}^{n} f(x_i,\theta)\, dx_1 \ldots dx_n$$

Proof Assume that X is a continuous r.v.

$$E(Y) = k(\theta) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} u(x_1, x_2, \ldots x_n) \prod_{i=1}^{n} f(x_i;\theta)\, dx_1\, dx_2 \ldots dx_n$$

Differentiating w.r.t. θ (this requires strong conditions on u, f)

$$k'(\theta) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} u(x_1, x_2, \ldots, x_n)\left(\sum_{i=1}^{n} \frac{1}{f(x_i;\theta)} \frac{\partial f(x_i;\theta)}{\partial\theta}\right) \prod_{i=1}^{n} f(x_i,\theta)\, dx_1 \ldots dx_n$$

$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} u(x_1, \ldots x_n)\left(\sum_{i=1}^{n} \frac{\partial \log(x_i;\theta)}{\partial\theta}\right) \prod_{i=1}^{n} f(x_i;\theta)\, dx_1 \ldots dx_n$$

Define a random variable $Z = \sum_{i=1}^{n} \frac{\partial \log f(X_i;\theta)}{\partial \theta}$

then $E[Z] = 0$.

and $Var(Z) = nI$.

Define a random variable $Z = \sum_{i=1}^{n} \frac{\partial \log f(x_i;\theta)}{\partial \theta}$

then $E[Z] = 0$.

and $Var(Z) = nI$.

Also, $k'(\theta) = E(YZ)$.

Therefore, $k'(\theta) = E(YZ) = E(Y)E(Z)$

$+ \rho \sigma_y \sqrt{nI(\theta)}$

where ρ is the correlation coefficient between Y and Z.

Define a random variable $Z = \sum_{i=1}^{n} \frac{\partial \log f(x_i;\theta)}{\partial \theta}$

then $E[Z] = 0.$

and $Var(Z) = nI.$

since
$Cov(Y,Z)$
$= E(YZ) -$
$E(Y)E(Z)$

Also, $k'(\theta) = E(YZ).$

Therefore, $k'(\theta) = E(YZ) = E(Y)E(Z)$
$+ \rho\, \sigma_y \sqrt{nI(\theta)}$

where ρ is the correlation coefficient between Y and Z.

$\therefore \quad k'(\theta) = f\,\sigma_y \sqrt{nI(\theta)}$

or $\quad f = \dfrac{k'(\theta)}{\sigma_y \sqrt{nI(\theta)}}$

Since $f^2 \leqslant 1$, we have

$$\frac{(k'(\theta))^2}{\sigma_y^2 \cdot nI(\theta)} \leq 1$$

$$\therefore \quad k'(\theta) = \rho\, \sigma_y \sqrt{n I(\theta)}$$

or $$\rho = \frac{k'(\theta)}{\sigma_y \sqrt{n I(\theta)}}$$

Since $\rho^2 \leq 1$, we have

$$\frac{(k'(\theta))^2}{\sigma_y^2 \cdot n I(\theta)} \leq 1$$

or $$\boxed{Var(Y) \geq \frac{(k'(\theta))^2}{n I(\theta)}}$$

As a consequence, if $Y = u(X_1, X_2, \dots, X_n)$ is an unbiased estimator of θ, then

$$k(\theta) = \theta \Rightarrow k'(\theta) = 1$$

As a consequence, if $Y = u(X_1, X_2, \dots, X_n)$ is an unbiased estimator of θ, then

$$k(\theta) = \theta \Rightarrow k'(\theta) = 1$$

In that case,

$$\boxed{Var(Y) \geq \frac{1}{nI(\theta)}}$$

As a consequence, if $Y = u(X_1, X_2, \ldots X_n)$ is an unbiased estimator of θ, then

$$k(\theta) = \theta \Rightarrow k'(\theta) = 1$$

In that case,

$$\boxed{Var(Y) \geq \frac{1}{n I(\theta)}}$$

Example. For the Bernoulli trials, $I(\theta) = \dfrac{1}{\theta(1-\theta)}$.

The m.l.e. of Bernoulli trial is $\overline{X}$.

The mean and variance of $\overline{X}$ are θ and $\dfrac{\theta(1-\theta)}{n}$ respectively. Therefore, in the case of Bernoulli trial, mle attains the lower bound.

1. **Efficient Estimator** Let Y be an unbiased estimator of a parameter θ in the case of point estimation. The estimator Y is called an efficient estimator of θ if the variance of Y attains the Rao-Cramer's lower bound.

2. **Efficiency** Whenever this is well-defined, the ratio of the Rao-Cramer's lower bound to the actual variance of any unbiased estimator of a parameter is called the efficiency of that estimator

Example. ① Poisson Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from the Poisson distribution with mean $\theta > 0$.

$\bar{X}$ is an m.l.e. of θ.

$$\frac{\partial \log f(x;\theta)}{\partial \theta} = \frac{\partial}{\partial \theta}\left(x \log \theta - \theta - \log x!\right)$$

$$= \frac{x}{\theta} - 1 = \frac{x-\theta}{\theta}$$

$$\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2} = \frac{-x}{\theta^2}$$

Example. (1) Poisson Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from the Poisson distribution with mean $\theta > 0$.

$\bar{X}$ is an m.l.e. of θ.

$$\frac{\partial \log f(x;\theta)}{\partial \theta} = \frac{\partial}{\partial \theta}\left(x \log \theta - \theta - \log x!\right)$$

$$= \frac{x}{\theta} - 1 = \frac{x-\theta}{\theta}$$

$$\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2} = \frac{-x}{\theta^2} \Rightarrow I(\theta) = -E\left[\frac{-X}{\theta^2}\right]$$

$$= \frac{\theta}{\theta^2} = 1/\theta.$$

Example. (1) Poisson Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from the Poisson distribution with mean $\theta > 0$.

$\bar{X}$ is an m.l.e. of θ.

the variance of $\bar{X}$ is θ/n,

$$\frac{\partial \log f(x;\theta)}{\partial \theta} = \frac{\partial}{\partial \theta}\left(x \log \theta - \theta - \log x!\right)$$

$$= \frac{x}{\theta} - 1 = \frac{x-\theta}{\theta}$$

$$\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2} = \frac{-x}{\theta^2} \Rightarrow I(\theta) = -E\left[\frac{-X}{\theta^2}\right]$$

$$= \frac{\theta}{\theta^2} = 1/\theta.$$

Example. ① Poisson Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from the Poisson distribution with mean $\theta > 0$.

$\bar{X}$ is an mle. of θ.

the variance of $\bar{X}$ is θ/n, so is the Rao-Cramer lower bound.

$$\frac{\partial \log f(x;\theta)}{\partial \theta} = \frac{\partial}{\partial \theta}\left(x \log \theta - \theta - \log x!\right)$$

$$= \frac{x}{\theta} - 1 = \frac{x-\theta}{\theta}$$

$$\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2} = \frac{-x}{\theta^2} \Rightarrow I(\theta) = -E\left[\frac{-X}{\theta^2}\right]$$

$$= \frac{\theta}{\theta^2} = 1/\theta.$$

Example. (1) Poisson Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from the Poisson distribution with mean $\theta > 0$.

$\boxed{\bar{X} \text{ is an m.l.e. of } \theta.}$

the variance of $\bar{X}$ is θ/n, so is the Rao-Cramer lower bound.

an efficient estima-
- tor.

$$\frac{\partial \log f(x;\theta)}{\partial \theta} = \frac{\partial}{\partial \theta}\left(x \log \theta - \theta - \log x!\right)$$

$$= \frac{x}{\theta} - 1 = \frac{x-\theta}{\theta}$$

$$\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2} = \frac{-x}{\theta^2} \Rightarrow I(\theta) = -E\left[\frac{-X}{\theta^2}\right]$$

$$= \frac{\theta}{\theta^2} = 1/\theta.$$

Example (2) Beta $(\theta, 1)$ distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n \geq 2$ from a distribution with p.d.f.

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1} & \text{for } 0<x<1 \\ 0 & \text{otherwise.} \end{cases}$$

$\boxed{\theta \in (0,\infty).}$ This is called the beta distribution with parameters θ and 1.

$$\log f(x;\theta) = \log\theta + (\theta-1)\log x.$$

$$\Rightarrow \frac{\partial \log f(x;\theta)}{\partial\theta} = \frac{1}{\theta} + \log x.$$

Example ② Beta $(\theta, 1)$ distribution

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1} & \text{for } 0<x<1 \\ 0 & \text{otherwise.} \end{cases}$$

$$\log f(x;\theta) = \log\theta + (\theta-1)\log x.$$

$$\Rightarrow \quad \frac{\partial \log f(x;\theta)}{\partial\theta} = \frac{1}{\theta} + \log x.$$

$$\therefore \quad \frac{\partial^2 \log f(x;\theta)}{\partial\theta^2} = -\frac{1}{\theta^2}$$

$$\therefore \quad I(\theta) = 1/\theta^2.$$

Let us find the mle of beta$(1,\theta)$.

$$l(\theta) = \log\left(\prod_{i=1}^{n} \theta x_i^{\theta-1}\right)$$

$$= \sum_{i=1}^{n} \log\left(\theta x_i^{\theta-1}\right)$$

$$= \sum_{i=1}^{n} \log\theta + \sum_{i=1}^{n} (\theta-1) \log x_i$$

$$\therefore \frac{\partial l(\theta)}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i$$

Setting this to zero, we get

$$\hat{\theta} = \frac{-n}{\sum_{i=1}^{n} \log X_i}.$$

$$l(\theta) = \log\left(\prod_{i=1}^{n} \theta x_i^{\theta-1}\right)$$

$$= \sum_{i=1}^{n} \log\left(\theta x_i^{\theta-1}\right)$$

$$= \sum_{i=1}^{n} \log\theta + \sum_{i=1}^{n} (\theta-1)\log x_i$$

$$\therefore \frac{\partial l(\theta)}{\partial\theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i$$

Setting this to zero, we get

$$\hat{\theta} = \frac{-n}{\sum_{i=1}^{n} \log X_i}$$

What is the distribution of $\hat{\theta}$?

Step 1 Let $Y_i = -\log X_i$

then

$$P(Y_i \leq y_i) = P(-\log X_i \leq y_i)$$

$$= P(\log X_i \geq -y_i)$$

$$= P(X_i \geq e^{-y_i})$$

$$= \int_{e^{-y_i}}^{1} \theta x^{\theta-1}\, dx$$

$$= \left[x^{\theta}\right]_{e^{-y_i}}^{1} = 1 - \left(e^{-y_i}\right)^{\theta}$$

$$= 1 - e^{-\theta y_i}$$

Step 1 Let $Y_i = -\log X_i$

then

$$P(Y_i \leq y_i) = 1 - e^{-\theta y_i}$$

$$\therefore f_Y(y, \theta) = \theta e^{-\theta y_i}$$

This is a Gamma (α, β) distribution with $\alpha = 1$, $\beta = 1/\theta$

Step 2

Recall that the mgf of Gamma dist

is $$M(t) = \frac{1}{(1-\beta t)^{\alpha}} \quad \text{for } t < 1/\beta$$

$$= \frac{1}{(1-t/\theta)} \quad \text{for } t < \theta.$$

Step 2 Recall that the mgf of Gamma dist

is $M(t) = \frac{1}{(1-\beta t)^{\alpha}}$ for $t < 1/\beta$

$$= \frac{1}{(1-t/\theta)} \text{ for } t < \theta.$$

Step 3 Let $W = \sum_{i=1}^{n} Y_i$ then by independence

$$M_W(t) = \frac{1}{(1-t/\theta)^n}$$

$\therefore W \sim \Gamma(n, 1/\theta)$.

Step 4

Show that

$$E[W^{-1}] = \frac{\theta}{n-1} \qquad (\text{Excercise !})$$

Step 5

Therefore

$$E[\hat{\theta}] = n\,E[W^{-1}] = \frac{n\theta}{n-1}.$$

This shows that mle for beta distribution $b(1,\theta)$ is not unbiased.

Step 4 Show that

$$E[W^{-1}] = \frac{\theta}{n-1} \qquad \text{(Excercise !)}$$

Step 5

Therefore

$$E[\hat{\theta}] = n E[W^{-1}] = \frac{n\theta}{n-1}.$$

This shows that mle for beta distribution $b(1,\theta)$ is not unbiased.

It follows that the estimator $\boxed{\frac{n-1}{n}\hat{\theta}}$ is unbiased.

Step 6 Show that

$$E[W^{-2}] = \frac{\theta^2}{(n-1)(n-2)} \quad \text{(Excercise!)}$$

Step 7 Therefore $E[\hat{\theta}^2] = \dfrac{n^2\theta^2}{(n-1)(n-2)}$

As a result,

$$\begin{aligned} Var(\hat{\theta}) &= E[\hat{\theta}^2] - E[\hat{\theta}]^2 \\ &= \frac{n^2\theta^2}{(n-1)(n-2)} - \frac{n^2\theta^2}{(n-1)^2} = \frac{\theta^2 n^2}{(n-1)^2(n-2)} \end{aligned}$$

Step 6 Show that

$$E[W^{-2}] = \frac{\theta^2}{(n-1)(n-2)} \quad \text{(Excercise!)}$$

Step 7 Therefore $E[\hat{\theta}^2] = \dfrac{n^2\theta^2}{(n-1)(n-2)}$

As a result,

$$Var(\hat{\theta}) = E[\hat{\theta}^2] - E[\hat{\theta}]^2$$

$$= \frac{n^2\theta^2}{(n-1)(n-2)} - \frac{n^2\theta^2}{(n-1)^2} = \frac{\theta^2 n^2}{(n-1)^2(n-2)}$$

Consequenty, $Var\left(\frac{n-1}{n}\hat{\theta}\right) = \dfrac{\theta^2}{(n-2)}$

We calculated earlier that

$I(\theta) = \theta^{-2}$ and hence the

Rao-Cramér lower bound is $\boxed{\dfrac{\theta^2}{n}}$.

We calculated earlier that

$I(\theta) = \theta^{-2}$ and hence the

Rao-Cramér lower bound is $\boxed{\dfrac{\theta^2}{n}}$.

The unbiased estimator $\dfrac{(n-1)\hat{\theta}}{n}$ is not efficient.

The efficiency is $\dfrac{n-2}{n}$. This converges to 1 as $n \longrightarrow \infty$.

We calculated earlier that

$I(\theta) = \theta^{-2}$ and hence the

Rao-Cramér lower bound is $\boxed{\dfrac{\theta^2}{n}}$.

The unbiased estimator $\dfrac{(n-1)\hat{\theta}}{n}$ is not efficient.

The efficiency is $\dfrac{n-2}{n}$. This converges to 1 as $n \longrightarrow \infty$. We will say that

$\dfrac{(n-1)\hat{\theta}}{n}$ is asymptotically efficient.

Theorem In addition to all the conditions on f, also assume that $f(x;\theta)$ is three times differentiable as a function of θ. Further for all $\theta \in \Omega$, $\exists$ c and a function $M(x)$ s.t.

$$\left| \frac{\partial^3 \log f(x;\theta)}{\partial^3 \theta} \right| \leq M(x) \quad \text{with}$$

$E_{\theta_0}[M(x)] < \infty$, for all $\theta_0 - c < \theta < \theta_0 + c$ and all x in the support of X. Finally assume that $0 < I(\theta_0) < \infty$.

Theorem Under suitable conditions on $f(x;\theta)$, any consistent sequence of solutions of mle equations satisfies

$$\boxed{\sqrt{n}\,(\hat{\theta}-\theta_0) \xrightarrow{D} N\left(0, \frac{1}{I(\theta_0)}\right)}$$

Sketch of Proof

Recall that

$$\ell(\theta) = \sum_{i=1}^{n} \log f(x_i;\theta)$$

and $$\ell'(\theta) = \sum_{i=1}^{n} \frac{\partial \log f(x_i;\theta)}{\partial \theta}.$$

Sketch of Proof

Recall that

$$l(\theta) = \sum_{i=1}^{n} \log f(x_i;\theta)$$

and

$$l'(\theta) = \sum_{i=1}^{n} \frac{\partial \log f(x_i;\theta)}{\partial \theta}.$$

Expanding $l'(\theta)$ into a Taylor series of order 2 about θ_0 and evaluating at $\hat{\theta}_n$, we have

$$l'(\hat{\theta}_n) = l'(\theta_0) + (\hat{\theta}_n - \theta_0)\, l''(\theta) + \frac{1}{2}(\hat{\theta}_n - \theta_0)^2 l'''(\theta_n^*)$$

where $\theta_n^* \in (\theta_0, \hat{\theta}_n)$.

Sketch of Proof

Recall that

$$\ell(\theta) = \sum_{i=1}^{n} \log f(x_i;\theta)$$

and $$\ell'(\theta) = \sum_{i=1}^{n} \frac{\partial \log f(x_i;\theta)}{\partial \theta}.$$

Expanding $\ell'(\theta)$ into a Taylor series of order 2 about θ_0 and evaluating at $\hat{\theta}_n$, we have

$$\ell'(\hat{\theta}_n) = \ell'(\theta_0) + (\hat{\theta}_n - \theta_0)\,\ell''(\theta) + \frac{1}{2}(\hat{\theta}_n - \theta_0)^2\,\ell'''(\theta_n^*)$$

where $\theta_n^* \in (\theta_0, \hat{\theta}_n)$.

This is zero since $\hat{\theta}_n$ solves the mle eqn.

Rearranging

$$l'(\hat{\theta}_n) = l'(\theta_0) + (\hat{\theta}_n - \theta_0)\, l''(\theta) + \frac{1}{2}(\hat{\theta}_n - \theta_0)^2\, l'''(\theta_n^*)$$

we get

Rearranging

$$l'(\hat{\theta}_n) = l'(\theta_0) + (\hat{\theta}_n - \theta_0)\, l''(\theta) + \frac{1}{2}(\hat{\theta}_n - \theta_0)^2\, l'''(\theta_n^*),$$

we get

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) = \frac{n^{-1/2}\, l'(\theta_0)}{-n^{-1}\, l''(\theta_0) - (2n)^{-1}(\hat{\theta}_n - \theta_0)\, l'''(\theta_0)}$$

Rearranging

$$\ell'(\hat{\theta}_n) = \ell'(\theta_0) + (\hat{\theta}_n - \theta_0)\,\ell''(\theta) + \frac{1}{2}(\hat{\theta}_n - \theta_0)^2\,\ell'''(\theta_n^*),$$

we get

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) = \frac{n^{-1/2}\,\ell'(\theta_0)}{-n^{-1}\,\ell''(\theta_0) - (2n)^{-1}(\hat{\theta}_n - \theta_0)\,\ell'''(\theta_0)}$$

By Central limit Theorem,

$$\frac{1}{\sqrt{n}}\,\ell'(\theta_0) = \frac{\sqrt{n}}{n}\sum_{i=1}^{n}\frac{\partial \log f(x_i,\theta)}{\partial\theta}$$

$$\xrightarrow{D} N(0, I(\theta_0))$$

Rearranging

$$l'(\hat{\theta}_n) = l'(\theta_0) + (\hat{\theta}_n - \theta_0)\, l''(\theta) + \frac{1}{2}(\hat{\theta}_n - \theta_0)^2 l'''(\theta_n^*),$$

we get

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) = \frac{n^{-1/2}\, l'(\theta_0)}{-n^{-1}\, l''(\theta_0) - (2n)^{-1}(\hat{\theta}_n - \theta_0)\, l'''(\theta_0)}$$

By Central Limit Theorem,

$$\frac{1}{\sqrt{n}}\, l'(\theta_0) = \frac{\sqrt{n}}{n} \sum_{i=1}^{n} \frac{\partial \log f(x_i, \theta)}{\partial \theta}$$

$$\xrightarrow{D} N(0, I(\theta_0))$$

var of $\frac{\partial \log f}{\partial \theta}$

Also, by the law of large numbers,

$$-\frac{1}{n} l''(\theta_0) = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log f(x_i;\theta_0)}{\partial \theta^2}$$

$$\xrightarrow{P} I(\theta_0)$$

Also, by the law of large numbers,

$$-\frac{1}{n} l''(\theta_0) = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log f(x_i;\theta_0)}{\partial \theta^2}$$

$$\xrightarrow{P} I(\theta_0)$$

Finally for the third expression

$\hat{\theta}_n - \theta_0 \xrightarrow{P} 0$, and the expression

$$\frac{l'''(\theta_0)}{n} \leq \frac{\sum M(x_i)}{n} \longrightarrow E_{\theta_0}(M(X))$$

(Therefore the expression $\frac{l'''(\theta_0)}{n}(\hat{\theta}_n - \theta_0) \xrightarrow{P} 0.$

as $n \to \infty$

Therefore,

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{D} \frac{N(0, I(\theta))}{I(\theta)}$$

$$= N\left(0, \frac{1}{I(\theta)}\right)$$

$\blacksquare$

Therefore,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \frac{N(0, I(\theta))}{I(\theta)}$$

$$= \boxed{N\left(0, \frac{1}{I(\theta_0)}\right)}$$

This immediately gives us an approximate confidence interval for mle.

$$\left(\hat{\theta}_n - z_{\alpha/2} \frac{1}{\sqrt{n I(\hat{\theta}_n)}} , \hat{\theta}_n + z_{\alpha/2} \frac{1}{\sqrt{n I(\hat{\theta}_n)}} \right)$$