ASYMPTOTIC DISTRIBUTION OF MAXIMUM LIKELIHOOD ESTIMATORS

1. INTRODUCTION

The statistician is often interested in the properties of different estimators. Rather than determining these properties for every estimator, it is often useful to determine properties for classes of estimators. For example it is possible to determine the properties for a whole class of estimators called extremum estimators. Members of this class would include maximum likelihood estimators, nonlinear least squares estimators and some general minimum distance estimators. Another class of estimators is the method of moments family of estimators. This section will derive the large sample properties of maximum likelihood estimators as an example. Additional discussion is contained in Amemiya [1], Gallant[3], Malinvaud[5], or Theil[7].

2. STANDARD ASSUMPTIONS FOR ASYMPTOTIC ANALYSIS

2.1. **Variables and Parameters.** Let X = (X_1 ... X_n) be a vector of random variables and $\theta = (\theta_1, \theta_2, \ldots \theta_k)$ a vector of unknown parameters where θ_0 is the true parameter vector. For a simple proof here, let θ be a scalar. The results generalize.

2.2. **Regularity Conditions.**

a: For almost all X, the derivatives

$$\frac{\partial log f}{\partial \theta}, \frac{\partial^2 log f}{\partial \theta^2} \text{ and } \frac{\partial^3 log f}{\partial \theta^3}$$

exist $\forall\theta$ belonging to a nondegenerate interval A

b: $\forall\theta\epsilon$ A, where

$$\left|\frac{\partial f}{\partial \theta}\right| < F_1(x), \left|\frac{\partial^2 f}{\partial \theta^2}\right| < F_2(x) \text{ and } \left|\frac{\partial^3 log f}{\partial \theta^3}\right| < H(x),$$

the functions F_1 and F_2 are integrable over (-∞,∞) while

$$\int_{-\infty}^{\infty} H(x) f(x,\theta) dx < M$$

where M is positive and independent of θ

c:

$$\forall\, \theta\, \varepsilon\, A, \int_{-\infty}^{\infty} \left(\frac{\partial log f}{\partial \theta}\right)^2 f(x,\theta) dx$$

is positive and finite i.e..

$$Var \left(\frac{\partial log f}{\partial \theta}\right)_{\theta_0}$$

is positive and finite

2.3. **Identities.** We will need some following identities from the lecture on the large sample theory and the Cramer-Rao lower bound.

Date: September 24, 2004.

2.3.1. *The likelihood function.* Consider a random sample $(X_1, \ldots, X_n)$ from some population characterized by the parameter θ and density function $f(x;\theta)$. The distribution is assumed to be continuous and so the joint density which is the same as the likelihood function is given by

$$L(x_1, x_2, \ldots, x_n) = f(x_1;\theta)f(x_2;\theta)\ldots f(x_n;\theta) \tag{1}$$

The following assumptions, called regularity conditions, are used to develop the Cramer-Rao lower bound.

$$\begin{aligned}
&\text{(i)} \quad f(\cdot;\theta) \text{ and } L(\cdot;\theta) \text{ are } C^2 \text{ w.r.t. } \theta.\\
&\text{(ii)} \quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} L(x_1, \ldots, x_n;\, \theta)\, dx_1\, dx_2 \ldots dx_n = 1.\\
&\text{(iii)} \quad \text{The limits of integration don't depend on } \theta.\\
&\text{(iv)} \quad \text{Differentiation under the integral sign is allowed.}
\end{aligned} \tag{2}$$

The notation C^2 means that the function is twice continuously differentiable. The regularity conditions imply the following theorem

Theorem 1. *If a likelihood function is regular then*

$$E\left[\frac{\partial log L(\cdot;\theta)}{\partial\theta_i}\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left[\frac{\partial}{\partial\theta}\, log\, L(x,\theta)\right] L(x,\theta)dx = 0 \tag{3}$$

Proof. Take the derivative of the condition in (ii) and then multiply and divide by $L(\cdot;\theta)$ inside the integral.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} L(x_1, \ldots, x_n;\, \theta)\, dx_1\, dx_2 \cdots\cdots dx_n = 1$$

$$\begin{aligned}
&\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} L(x_1, \ldots, x_n;\, \theta)\, dx_1 \ldots dx_n\\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\partial L(x_1, \ldots, x_n;\, \theta)}{\partial\theta}\, dx_1 \ldots dx_n = 0\\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\partial L(x_1, \ldots, x_n;\theta)}{\partial\theta}\frac{L(\cdot;\theta)}{L(\cdot;\theta)}\, dx_1 \ldots dx_n = 0\\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\partial \log L(x_1, \ldots, x_n;\, \theta)}{\partial\theta} L(\cdot;\theta)\, dx_1 \ldots dx_n = 0\\
&\Rightarrow E\left[\frac{\partial \log L(\cdot;\, \theta)}{\partial\theta}\right] = 0
\end{aligned} \tag{4}$$

□

In the case of a single parameter θ, the Fisher information number is given by

$$R(\theta) = -E\left[\frac{\partial^2 \log L(X,\theta)}{\partial\theta^2}\right] \tag{5}$$

We can show that this number is the variance of the derivative of the log likelihood function with respect to θ, i.e.,

$$\mathrm{Var}\left[\frac{\partial \log L(X;\theta)}{\partial \theta}\right]$$

This can be shown by differentiating the penultimate expression in equation 4 with respect to θ (using the product rule) and remembering that the expected value of the derivative of the log likelihood function is zero, so that its variance is just the expected value of its square.

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\partial \log L(x_1,...,x_n;\theta)}{\partial\theta}L(\cdot;\theta)\,dx_1\,\ldots\,dx_n = 0\\
&\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\partial \log L(\cdot;\theta)}{\partial\theta}L(\cdot;\theta)\,dx_1\,\ldots\,dx_n\\
&\quad=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\partial^2 \log L(\cdot;\theta)}{\partial\theta^2}L(\cdot;\theta)\,dx_1\,\ldots dx_n\\
&\quad+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\partial L(\cdot;\theta)}{\partial\theta}\frac{\partial \log L(\cdot;\theta)}{\partial\theta}\,dx_1\,\ldots\,dx_n = 0\\
&\quad=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\partial^2 log L(\cdot;\theta)}{\partial\theta^2}L(\cdot;\theta)\,dx_1\,\ldots\,dx_n\\
&\quad+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\partial \log L(\cdot;\theta)}{\partial\theta}\frac{\partial \log L(\cdot;\theta)}{\partial\theta}L(\cdot;\theta)\,dx_1\,\ldots\,dx_n = 0\\
&\quad= E\left[\frac{\partial^2 log L(\cdot;\theta)}{\partial\theta^2}\right]+\mathrm{Var}\left[\frac{\partial log L(X;\theta)}{\partial\theta}\right]=0\\
&\quad\Rightarrow \mathrm{V}\left[\frac{\partial log L(X;\theta)}{\partial\theta}\right]=-E\left[\frac{\partial^2 \log L(\cdot;\theta)}{\partial\theta^2}\right]
\end{aligned} \tag{6}$$

3. Consistency of MLE

For a random sample, X = (X_1 ... X_n), the likelihood function is product of the individual density functions and the log likelihood function is the sum of the individual likelihood functions, i.e.,

$$log\,L\,(X,\,\theta)\;=\;\sum_{i=1}^{n}log\,f\,(x_i\,\theta\,) \tag{7}$$

Let θ_0 be the unknown true parameter such that θ is in the interval A, i.e., $\theta\,\epsilon$ int(A). To find the maximum likelihood estimators we solve the equations:

$$\frac{\partial log L(X,\theta)}{\partial\theta}\;=\;0 \tag{8}$$

where

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

Apply a Taylor Series expansion to

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

for some $\theta \epsilon$ A

$$\frac{\partial log\, f\,(X_i\,,\,\theta)}{\partial\,\theta} \;=\; \left(\frac{\partial\, log\, f\,(\,X_i\,,\,\theta\,)}{\partial\,\theta}\right)_{\theta_0} + (\theta - \theta_0\,)\,\left(\frac{\partial^2\, log\, f\,(X_i\,,\,\theta\,)}{\partial\,\theta^2}\right)_{\theta_0} + \frac{1}{2}\,(\theta\;-\;\theta_0\,)^{\,2}\,\left(\frac{\partial^3\, log\, f\,(X_i\,,\,\theta\,)}{\partial\,\theta^3}\right)_{\theta^*} \tag{9}$$

where θ^* is between θ_0 and θ, and $\theta^* \in$ A. This is equal to the following

$$\frac{\partial\, log\, f\,(X_i\,,\,\theta\,)}{\partial\,\theta} \;=\; \left(\frac{\partial\, log\, f\,(X_i\,,\,\theta\,)}{\partial\,\theta}\right)_{\theta_0} +\;(\;\theta\;-\;\theta_0)\,\left(\frac{\partial^2\, log\, f\,(X_i\,,\,\theta\,)}{\partial\,\theta^2}\right)_{\theta_0} + \frac{1}{2}\,(\,\theta\;-\;\theta_0\,)^2\,\zeta\,H\,(\,X_i\,) \tag{10}$$

for some $|\zeta| < 1$.

Now define T as follows

$$T \;=\; \frac{1}{n}\frac{\partial\, log\, L\,(X\,,\,\theta\,)}{\partial\,\theta} \;=\; \beta_0\,(\,X\,) \;+\; (\;\theta\;-\;\theta_0\,)\,\beta_1\,(\,X\,) + \frac{1}{2}\,\zeta\;(\;\theta\;-\;\theta_0\,)^{\,2}\;\;\beta_2\,(\,X\,) \;=\; 0 \tag{11}$$

where

$$\beta_0\,(\,X\,) \;=\; \frac{1}{n}\,\sum_{i=1}^{n}\,\left(\frac{\partial\, log\, f\,(X_i\,,\,\theta)}{\partial\,\theta}\right)_{\theta_0} \tag{12}$$

and

$$\beta_1\,(\,X\,) \;=\; \frac{1}{n}\,\sum_{i=1}^{n}\,\left(\frac{\partial^2 log\, f\,(\,X_i\,,\,\theta)}{\partial\,\theta^2}\right)_{\theta_0} \tag{13}$$

and

$$\beta_2\,(\,X\,) \;=\; \frac{1}{n}\,\sum_{i=1}^{n}\,H(\,X_i\,) \tag{14}$$

We need to show that T has a root θ between limits $\theta_0 \pm \delta$ with a probability tending to 1, no matter how small the quantity δ.

Remember from Khintchine's Theorem that the sum of n independent and identically distributed (iid) random variables divided by n converges in probability to their expectations.

Theorem 2 (Khintchine)**.** *Let $X_1,\, X_2,\, \ldots$ be independent and identically distributed random variables with $E(X_i) = \mu < \infty$. Then*

$$\frac{1}{n}\sum_{t=1}^{n} X_t = \bar{X}_n \xrightarrow{P} \mu \tag{15}$$

Proof: Rao [6, p. 113].

So β_0 converges in probability to zero because

$$\beta_0\,(\,X\,) = \;\frac{1}{n}\,\sum_{i=1}^{n}\,\left(\frac{\partial\, log\, f\,(X_i\,,\,\theta\,)}{\partial\,\theta}\right)_{\theta_0} \tag{16}$$

and

$$E\left[\frac{\partial \log f\,(X_i\,,\,\theta)}{\partial\theta}\right]_{\theta_0} = \int \left(\frac{\partial \log f\,(X_i\,,\,\theta)}{\partial\,\theta}\right)_{\theta_0} f\,(x\,,\,\theta_0\,)\,dx \,=\, 0 \tag{17}$$

by equation 3 where we take n = 1 so f( ) = L( ).

Now let

$$E\left(\frac{\partial^2 \log f\,(X\,,\,\theta)}{\partial\,\theta^2}\right)_{\theta_0} = -k^2 \tag{18}$$

This is negative by the second order conditions for a maximum. So β_1(X) converges to -k^2 where k^2 is equal to

$$k^2 \,=\, -\int \left(\frac{\partial^2 \log f\,(X\,,\,\theta\,)}{\partial\,\theta^2}\right)_{\theta_0} f\,(\,x\,,\,\theta_0\,)\,dx \,=\, \int \left(\frac{\partial \log f(\,X\,,\,\theta)}{\partial\,\theta}\right)^2_{\theta_0} f\,(\,x\,,\,\theta_0\,)dx \tag{19}$$

where the second equality is based on equation 8 where n=1 and L( ) = f( ), and the zero expected value of

$$\frac{\partial \log f}{\partial\,\theta}$$

in equation 17. No matter which X_i is considered, H(X_i) has expectation between zero and M so it can be written as a nonnegative fraction of M i.e., γM where $|\gamma| < 1$. Specifically then

$$\begin{aligned} plim_{n\,\to\,\infty}\,\beta_0\,(\,X\,) \,&=0 \\ plim_{n\,\to\,\infty}\,\beta_1\,(\,X\,) \,&=\,-\,k^2 \\ plim_{n\,\to\,\infty}\,\beta_2\,(\,X\,) \,&=\gamma\,M \end{aligned} \tag{20}$$

Now let δ and ε be arbitrarily small positive numbers and let P(S) denote the joint density function of the random variables (X_1 ... X_n). For n sufficiently large say n > n_0 = n_0(δ, ε) we have

$$\begin{aligned} P_1 \,&=P(|\beta_0|\,\geq\,\delta^2)\,<\,\frac{1}{3}\varepsilon \\ P_2 \,&=P(\beta_1\,\geq\,-\frac{1}{2}\,k^2)\,<\,\frac{1}{3}\varepsilon \\ P_3 \,&=P(|\beta_2|\,\geq\,2M)\,<\,\frac{1}{3}\varepsilon \end{aligned} \tag{21}$$

Now let S denote the set of points where all three inequalities

$$|\beta_0|\,<\,\delta^2,\,\beta_1\,<\,-\frac{1}{2}\,k^2\,|\beta_2|\,<\,2M \tag{22}$$

are satisfied.

The complement to S denoted by S* consists of all points X = (X_1 ... X_n) such that at least one of the three inequalities is not satisfied. Thus we have P(S*) $\leq$ P_1 + P_2 + P_3 ¡ ε because P(S_1 + S_2 + ... S_n) $\leq$ P(S_1) + P(S_2) + ... Therefore P(S) > 1 - ε. Thus the probability that X = (X_1 ... X_n) belongs to the set S is greater than 1-ε as soon as n > n_0(δ,ε). Now let $\theta = \theta_0 \pm \delta$ and substitute into 11 so that

$$T \,=\, \beta_0(X) \,\pm\, \beta_1\,\delta \,+\, \frac{1}{2}\,\zeta\beta_2\,\delta^2 \tag{23}$$

Now consider that for points in S, $|\beta_0| < \delta^2$ and $|1/2\zeta\beta_2| <$ M because $|\zeta|$ is less than 1. This implies that $|1/2\zeta\,\beta_2\,\delta^2| <$ M δ^2, so that for every point X that is in the set S, the sum of the first and third terms is smaller in absolute value than δ^2 + Mδ^2 = [(M+1)δ^2]. Specifically,

$$|\beta_0 + \frac{1}{2}\zeta\beta_2\delta^2| < (M+1)\delta^2$$

Furthermore for points in S, $\beta_1\,\delta <$ - $\frac{1}{2}k^2\,\delta$. So if (M+1)δ^2 is smaller than the absolute value of $\frac{1}{2}$k$^2\delta$, or equivalently

$$\delta < \left|\frac{\frac{1}{2}k^2}{M+1}\right|$$

then the sign of the whole expression will be determined by the second term, so that we have

$$\frac{\partial \log L}{\partial\theta} < 0,\ \theta = \theta_0 + \delta$$
$$\frac{\partial \log L}{\partial\theta} > 0,\ \theta = \theta_0 - \delta$$

becuase β_1 tends to a negative number -k^2. Now remember the function is continuous at almost all X by regularity condition a. Thus for arbitrarily small δ and ε the likelihood equation will (with a probability exceeding 1-ε) have a root between the limits $\theta_0 \pm \delta$ as soon as n¿n$_0$ (δ,ε).

4. Asymptotic Normality of MLE

Now let

$$\hat{\theta} = \hat{\theta}\,(X_1\,,\,X_2\,,\,\cdots\,,\,X_n\,)$$

be the solution to the likelihood equation (8) above. Now reformulate equation 11 as

$$\left(\hat{\theta} - \theta_0\right)\left[\beta_1\,(\,X\,) + \frac{1}{2}\zeta\,\left(\hat{\theta} - \theta_0\right)\,\beta_2(\,X\,)\right] = -\beta_0\,(\,X\,) \tag{24}$$

Now rearrange the expression as follows

$$\begin{aligned}
(\hat{\theta} - \theta_0)\left[\beta_1(X) + \frac{1}{2}\zeta(\hat{\theta} - \theta_0)\,\beta_2(\,X\,)\right] &= -\,\beta_0\,(X\,) \\
\Rightarrow\ (\,\hat{\theta} - \theta_0)\left[\beta_1(X\,) + \frac{1}{2}\,\zeta(\hat{\theta} - \theta_0)\,\beta_2(X\,)\right] &= -\,\frac{1}{n}\Sigma_{i=1}^{n}\left(\frac{\partial\,log\,f\,(X_i\,,\,\theta\,)}{\partial\,\theta}\right)_{\theta_0} \\
\Rightarrow\ \sqrt{n}\left(\hat{\theta} - \theta_0\right)\left[-\beta_1(X\,) - \frac{1}{2}\,\zeta(\hat{\theta} - \theta_0)\,\beta_2(X\,)\right] &= \frac{1}{\sqrt{n}}\,\Sigma_{i=1}^{n}\left(\frac{\partial\,log\,f\,(X_i\,,\,\theta)}{\partial\,\theta}\right)_{\theta_0} \\
\Rightarrow\ k\,\sqrt{n}\left(\hat{\theta} - \theta_0\right) &= \frac{\frac{1}{k\,\sqrt{n}}\,\sum_{i=1}^{n}\left(\frac{\partial\,log\,f\,(X_i,\,\theta\,)}{\partial\,\theta}\right)_{\theta_0}}{\frac{-\beta_1\,(X\,)}{k^2}\ -\ \frac{\frac{1}{2}\,\zeta\left(\hat{\theta} - \theta_0\right)\beta_2(X)}{k^2}}
\end{aligned} \tag{25}$$

Now note that

$$\frac{-\beta_1(X)}{k^2} \xrightarrow{P} 1$$

because plim β_1(X) = -k^2. Further note that β_2(X) converges to a finite number γM and that

$$\left(\hat{\theta} - \theta_0\right)$$

converges in probability to zero so

$$\frac{\frac{1}{2}\zeta(\hat{\theta} - \theta_0)\beta_2(X)}{k^2}$$

converges to zero. The whole denominator of the fraction converges to 1 as n $\rightarrow \infty$. Note further that

$$\left(\frac{\partial \log f(X_i, \theta)}{\partial \theta}\right)_{\theta_0}$$

has mean zero by equation 17and has variance k^2 by equations 6 and 18. Given this information about $\left(\frac{\partial \log f(X_i, \theta)}{\partial \theta}\right)_{\theta_0}$ we can use the Lindberg-Levy central limit theorem to find the asymptotic distribution of a maximum likelihood estimator.

Theorem 3 (Lindberg-Levy Central Limit Theorem)**.** *Let $X_1, X_2, \ldots X_n$ be a sequence of independent and identically distributed random variables with finite mean μ and finite variance σ^2. Then the random variable*

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} N(0, \sigma^2) \tag{26a}$$

or

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{x_i - \mu}{\sigma} \xrightarrow{d} N(0, 1) \tag{26b}$$

We sometimes say that if

$$\sqrt{n}(\bar{x}_n - \mu) \xrightarrow{d} N(0, \sigma^2) \tag{27a}$$

then asymptotically

$$\bar{x}_n \sim N\left[\mu, \frac{\sigma^2}{n}\right] \tag{27b}$$

or

$$\bar{x}_n \xrightarrow{a} N\left[\mu, \frac{\sigma^2}{n}\right] \tag{27c}$$

In general for a vector of parameters θ with finite mean vector μ and covariance matrix Σ, the following holds

$$\begin{aligned} \sqrt{n}(\bar{\theta} - \mu) &\xrightarrow{d} N(0, \Sigma) \\ \bar{\theta} &\xrightarrow{a} N\left[\mu, \frac{1}{n}\Sigma\right] \end{aligned} \tag{28}$$

We say that $\bar{\theta}$ is asymptotically normally distributed with mean vector μ and covariance matrix $(1/n)\Sigma$.

Proof: Rao [6, p. 127] or Theil [7, pp. 368-9].

Applying the Lindberg-Levy central limit theorem to $\left(\frac{\partial \log f(X_i, \theta)}{\partial \theta}\right)_{\theta_0}$, it is clear that

$$\sum_{i=1}^{n} \left(\frac{\partial \, log \, f \, (X_i \, , \, \theta)}{\partial \, \theta} \right)_{\theta_0} \tag{29}$$

is asymptotically normal with mean zero and variance k^2n. Given the result in equation 29, the numerator in equation 25 is asymptotically normal with mean zero and variance $(1/k^2n) \cdot (k^2n) = 1$. By the Slutsky theorem on convergence, the right hand ratio in equation 25 converges to a standard normal variable. Thus $\hat{\theta}$ is asymptotically normal with mean θ_0 and variance $\frac{1}{k^2 \, n}$, i.e.,

$$\begin{aligned} \hat{\theta} &\xrightarrow{a} N \left[\theta_0, \, \frac{1}{k^2 \, n} \right] \\ &\xrightarrow{a} N \left[(\theta_0, \, -\frac{1}{E \left(\frac{\partial^2 \, log \, f \, (X \, , \theta)}{\partial \, \theta^2} \right)_{\theta_0}} \right] \end{aligned} \tag{30}$$

by equation 18.

5. Asymptotic Efficiency

To show efficiency we can show that this variance ($\frac{1}{k^2 \, n}$) is equal to the Cramer-Rao lower bound. Write this variance in terms of log f and then manipulate to obtain

$$\begin{aligned} Var \left(\Sigma_{i=1}^{n} \left(\frac{\partial \, log \, f(X_i \, , \, \theta)}{\partial \, \theta} \right) \right) \, &= \, k^2 \, n \\ \Rightarrow \frac{1}{k^2 \, n} \, &= \left(Var \, \left(\Sigma_{i=1}^{n} \, \left(\frac{\partial \, log \, f \, (X_i \, , \theta)}{\partial \, \theta} \right) \right) \right)^{-1} \\ &= \left(Var \, \left(\frac{\partial \, log \, L \, (X \, , \, \theta)}{\partial \, \theta} \right) \right)^{-1} \\ &= - \left(E \, \left[\frac{\partial^2 log L(X, \theta)}{\partial \theta^2} \right] \right)^{-1} \\ &= \frac{1}{\left(-E \left[\frac{\partial^2 \, log \, L \, (X \, , \, \theta)}{\partial \, \theta^2} \right] \right)} \end{aligned} \tag{31}$$

which is the lower bound.

6. Generalization

The MLE estimator of the vector θ_0, $\hat{\theta}_n$ is consistent and is asymptotically normally distributed with mean vector θ_0 and covariance matrix $R(\theta_0)^{-1}$ where $R(\theta_0)$ is the information matrix. A consistent estimator of $nR(\theta_0)^{-1}$ is $nR(\hat{\theta}_n)^{-1}$ Note that the likelihood equation may have more than one solution. If this is the case, they are asymptotically equivalent. If both are consistent, they are the same. The global maximum of the likelihood function provides a consistent estimator with unit probability.

REFERENCES

[1] Amemiya, T. *Advanced Econometrics.* Cambridge: Harvard University Press, 1985.
[2] Cramer, H. *Mathematical Methods of Statistics.* Princeton: Princeton University Press, 1946.
[3] Gallant, A.R. *Nonlinear Statistical Methods*. New York: Wiley, 1987.
[4] Goldberger, A.S. *Econometric Theory.* New York: Wiley, 1964.
[5] Malinvaud, E. *Statistical Methods of Econometrics.* Amsterdam: North-Holland, 1980.
[6] Rao, C.R. *Linear Statistical Inference and its Applications.* 2nd edition. New York: Wiley, 1973.
[7] Theil, H. *Principles of Econometrics*. New York: Wiley, 1971.