LECTURE 10

§16. EFFECTIVE ESTIMATORS. Rao-Cramer inequality

§16.1. THE RAO-CRAMER INEQUALITY

The Rao-Cramer inequality provides a lower bound for the variance of an estimator of a parameter. To prove this inequality we need the following auxiliary result.

Lemma 16.1. If random variables η_1 and η_2 have the second moments

$$E\eta_1^2, \qquad E\eta_2^2,$$

then

$$[E(\eta_1\,\eta_2)]^2 \le E\eta_1^2\,E\eta_2^2.$$

Proof: Since

$$E[(\lambda\,\eta_1 - \eta_2)^2]$$

is nonnegative for any real λ, that is

$$E[\lambda^2\,\eta_1^2 - 2\,\lambda\,\eta_1\,\eta_2 + \eta_2^2] \ge 0$$

and using the property of expectation we obtain

$$\lambda^2\,E[\eta_1^2] - 2\,\lambda\,E(\eta_1\,\eta_2) + E[\eta_2^2] \ge 0, \qquad \lambda \in \mathbf{R}.$$

This quadratic trinomial with respect to λ is nonnegative, so its discriminant is not positive:

$$[E(\eta_1\,\eta_2)]^2 \le E\eta_1^2\,E\eta_2^2.$$

The proof is complete.

This inequality is equality if and only if

$$\eta_2 = \lambda\,\eta_1$$

that is the random variables η_1 and η_2 differ by a constant factor.

Now we can prove the Rao-Cramer inequality.

Theorem. (Rao–Cramer Inequality). Let $f(x/\theta)$ be a probability density with continuous parameter θ. Let X_1, X_2,...,X_n be independent random variables with density $f(x/\theta)$, and let $\overline{\theta}(X_1, X_2, ..., X_n)$ be an estimator of θ. Assume that $f(x/\theta)$ satisfies two conditions for unknown parameter θ

a) We have

$$\frac{\partial}{\partial\theta}\left[\int ... \int \overline{\theta}(x_1, x_2, ..., x_n) \prod_{i=1}^{n} f(x_i/\theta)\, dx_1...dx_n\right] =$$

$$= \int ... \int \overline{\theta}(x_1, x_2, ..., x_n) \frac{\partial}{\partial\theta}\left[\prod_{i=1}^{n} f(x_i/\theta)\right] dx_1...dx_n.$$

b) For each θ, the variance of $\overline{\theta}(X_1, X_2, ..., X_n)$ is finite.

Then

$$\mathrm{Var}_\theta(\overline{\theta}(X_1, X_2, ..., X_n)) \geq \frac{(1 + b'(\theta))^2}{E[(L'(X_1, X_2, ..., X_n/\theta))^2]},$$

where E stands for expectation with respect to probability density function $f(x/\theta)$, and

$$b(\theta) = E(\overline{\theta}(X_1, X_2, ..., X_n)) - \theta,$$

and

$$L(X_1, X_2, ..., X_n/\theta)$$

is the logarithm of likelihood function.

Let us assume that distribution of η random variable contains unknown parameter θ. We perform n independent trials and obtain a sample $(X_1, X_2, ..., X_n)$.

There exists a method of construction of estimator which is called MLE (Maximal likelihood method). The MLE method states that the estimate $\bar{\theta}(X_1, X_2, ..., X_n)$ for unknown parameter θ is such estimate for which likelihood function

$$P(X_1, X_2, ..., X_n/\theta) = f(X_1/\theta)f(X_2/\theta)...f(X_n/\theta)$$

takes on a maximal value. Here $f(x/\theta)$ is either density function or probability $P(\eta = x/\theta)$ for discrete random variable.

Now assume that $\bar{\theta}(X_1, X_2, ..., X_n)$ ia an estimator of unknown parameter θ. Denote by

$$\sigma^2_{\bar{\theta}} = E\left[(\bar{\theta} - E(\bar{\theta}))^2\right]$$

the variance of $\bar{\theta}$ and by

$$b(\theta) = E(\bar{\theta}) - \theta$$

the deviation.

We consider the function

$$L(x_1, ..., x_n/\theta) = \ln P(x_1, ..., x_n/\theta)$$

and its derivative by θ

$$L'(x_1, ..., x_n/\theta) = \frac{P'(x_1, ..., x_n/\theta)}{P(x_1, ..., x_n/\theta)}$$

The following equalities is hold

$$\theta + b(\theta) = E(\overline{\theta}) = \int \overline{\theta}(x_1, ..., x_n)\, P(x_1, ..., x_n/\theta)\, dx_1\, dx_2 \ldots dx_n \tag{16.1}$$

and

$$1 = \int P(x_1, ..., x_n/\theta)\, dx_1\, dx_2 \ldots dx_n \tag{16.2}$$

Now we formally calculate the first derivative of (16.1) and (16.2) in θ. We get

$$1 + b'(\theta) = \int \overline{\theta}(x_1, ..., x_n)\, P'(x_1, ..., x_n/\theta)\, dx_1\, dx_2 \ldots dx_n$$

$$0 = \int P'(x_1, ..., x_n/\theta) dx_1\, dx_2 \ldots dx_n$$

or

$$1 + b'(\overline{\theta}) = \int \overline{\theta}(x_1, ..., x_n)\, \frac{P'(x_1, ..., x_n/\theta)}{P(x_1, ..., x_n/\theta)} P(x_1, ..., x_n/\theta)\, dx_1\, dx_2 \ldots dx_n =$$

$$= \int \overline{\theta}(x_1, ..., x_n)\, L'(x_1, ..., x_n/\theta) P(x_1, ..., x_n/\theta)\, dx_1\, dx_2 \ldots dx_n$$

$$0 = \int \frac{P'(x_1, ..., x_n/\theta)}{P(x_1, ..., x_n/\theta)} P(x_1, ..., x_n/\theta) dx_1 \ldots dx_n = \int L'(x_1, ..., x_n/\theta) P(x_1, ..., x_n/\theta) dx_1\, dx_2 \ldots dx_n$$

We rewrite these equalities in the following form

$$1 + b'(\theta) = E[\overline{\theta}(x_1, ..., x_n)\, L'(x_1, ..., x_n/\theta)]$$

$$0 = E[L'(x_1, ..., x_n/\theta)].$$

Multiplying the second equation by $E(\overline{\theta})$ and consider difference between the first equality and the second one, we obtain

$$1 + b'(\theta) = E[\overline{\theta}(x_1, ..., x_n) \, L'(x_1, ..., x_n/\theta)] - E[L'(x_1, ..., x_n/\theta) \, E[\overline{\theta}]],$$

or

$$1 + b'(\overline{\theta}) = E[L'(x_1, ..., x_n/\theta) \, [\overline{\theta} - E(\overline{\theta})]]$$

Two sides of equality take a square and apply lemma 16.1, we get

$$[1 + b'(\theta)]^2 \leq E[L'(x_1, ..., x_n/\theta)]^2 \, E[(\overline{\theta} - E(\overline{\theta})^2]$$

or

$$[1 + b'(\theta)]^2 \leq \sigma^2_{\overline{\theta}} \cdot E[L'(x_1, ..., x_n/\theta)]^2$$

Assuming, that

$$I(\theta) = E[L'(x_1, ..., x_n/\theta)]^2 > 0,$$

we have

$$\sigma^2_{\overline{\theta}} \geq \frac{[1 + b'(\theta)]^2}{I(\theta)} : \tag{16.3}$$

Thus, we see that the minimal value of the variance of estimate θ

$$\sigma^2_{\overline{\theta}}(min) = \frac{[1 + b'(\theta)]^2}{I(\theta)} :$$

In particular case, $(b(\theta) = 0)$ the variance of the effective estimate is

$$\sigma^2_{\overline{\theta}}(min) = \frac{1}{I(\theta)} :$$

which is a Corollary of the so-called Rao-Cramer inequality. This inequality gives a lower bound for the mean squared error for the unknown parameter θ that have the same bias function $b(\theta)$.

The last inequality is the Rao-Cramer inequality, and $I(\theta)$ is a Fisher information.

We have to find a criterion, by which we can find the estimate of unknown parameter with minimal variance.

Theorem 16.1. In order to estimate $\overline{\theta}$ for a given unknown parameter θ to be an estimate with minimal variance if and only if the likelihood function has the following form:

$$P(x_1, x_2, ..., x_n/\theta) = h(x_1, x_2, ..., x_n) \cdot \exp\{A(\theta) \cdot \overline{\theta}(x_1, ..., x_n) + B(\theta)\}, \tag{16.4}$$

where A and B depend only on θ, and $h(x_1, x_2, ..., x_n)$ depends solely on the values of our sample $(x_1, x_2, ..., x_n)$.

Proof: If the estimate $\overline{\theta}$ with minimal variance, then $L'(x_1, ..., x_n/\theta)$ and $\overline{\theta}(x_1, ..., x_n) - E(\overline{\theta})$ random variables differ from each other by not random factor

$$L'(x_1, ..., x_n/\theta) = \lambda(\theta)[\overline{\theta}(x_1, ..., x_n) - E(\overline{\theta})] = \lambda(\theta) \cdot \overline{\theta}(x_1, ..., x_n) - \lambda(\theta) \cdot E(\overline{\theta}),$$

where $\lambda(\theta)$ depends only on θ. Integrating this equality by θ, we get

$$L(x_1, ..., x_n/\theta) = A(\theta) \cdot \overline{\theta}(x_1, ..., x_n) + B(\theta) + C(x_1, ..., x_n)$$

since $\overline{\theta}$ does not depend on θ, and $E(\overline{\theta})$ depends only on θ. Here $A(\theta)$ and $B(\theta)$ depend only on θ, and $C(x_1, ..., x_n)$ depends only on sample $(x_1, ..., x_n)$.

Getting to likelihood function, we have

$$P(x_1, ..., x_n/\theta) = h(x_1, ..., x_n) \cdot \exp\{A(\theta) \cdot \overline{\theta}(x_1, ..., x_n) + B(\theta)\}.$$

Now prove the converse statement, that is if we have

$$P(x_1, ..., x_n/\theta) = h(x_1, ..., x_n) \cdot \exp\{A(\theta) \cdot \overline{\theta}(x_1, ..., x_n) + B(\theta)\}.$$

then

$$L(x_1, ..., x_n/\theta) = \ln h(x_1, ..., x_n) + A(\theta) \cdot \overline{\theta}(x_1, ..., x_n) + B(\theta)$$

and

$$L'(x_1, ..., x_n/\theta) = A'(\theta) \cdot \overline{\theta}(x_1, ..., x_n) + B'(\theta).$$

Calculate the expectation, we get

$$A'(\theta) \cdot E(\overline{\theta}) + B'(\theta) = 0,$$

therefore

$$B'(\theta) = -A'(\theta) \cdot E(\overline{\theta})$$

and

$$L'(x_1, ..., x_n/\theta) = A'(\theta) \, [\overline{\theta}(x_1, ..., x_n) - E(\overline{\theta})].$$

Since random variables $L'(x_1, ..., x_n/\theta)$ and $\overline{\theta}(x_1, ..., x_n) - E(\overline{\theta})$ differ from each other by not random factor, then by lemma 16.1, we get

$$[E[L'(x_1, ..., x_n/\theta) \cdot (\overline{\theta}(x_1, ..., x_n) - E(\overline{\theta}))]]^2 = E[L'(x_1, ..., x_n/\theta)]^2 \cdot E(\overline{\theta}(x_1, ..., x_n) - E(\overline{\theta})]^2].$$

Therefore we have

$$[1 + b'(\theta)]^2 = E[L'(x_1, ..., x_n/\theta)]^2 \, \sigma^2_{\overline{\theta}},$$

implying that

$$\sigma^2_{\overline{\theta}} = \frac{[1 + b'(\theta)]^2}{E[L'(x_1, ..., x_n/\theta)]^2}$$

hence $\overline{\theta}(x_1, ..., x_n)$ is an estimate with minimal variance for θ.

Example 16.1. Let us show that for normal distribution the estimate

$$\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$$

is an effective estimate for unknown expectation a. For simplicity we assume, that $\sigma^2 = 1$. The density function is

$$p(x/a) = \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{(x-a)^2}{2} \right\}.$$

The likelihood function is

$$P(x_1, ..., x_n/a) = \frac{1}{(2\pi)^{n/2}} \exp\left\{ -\frac{1}{2} \sum_{k=1}^n (x_k - a)^2 \right\}.$$

The likelihood function we can write in the following form

$$P(x_1, ..., x_n/a) = \frac{1}{(2\pi)^{n/2}} \exp\left\{ -\frac{1}{2} \sum_{k=1}^n x_k^2 \right\} \cdot \exp\{na\overline{X} - \frac{n}{2}a^2\},$$

here $A(a) = na,\ B(a) = -\dfrac{n}{2}a^2$ and

$$h(x_1, ..., x_n) = \frac{1}{(2\pi)^{n/2}} \exp\left\{ -\frac{1}{2} \sum_{k=1}^n x_k^2 \right\}$$

Therefore $\overline{X}$ is the effective estimate for parameter a.

Example 16.2. Probability $P(A) = p$ of the event A in n independent trials is unknown. If the event A is realized in m trials, we know that $\frac{m}{n}$ is the unbased estimate for p. We show, that it is an effective estimate for unknown p. Here the likelihood function is

$$P(m/p) = \binom{n}{m} p^m (1-p)^{n-m},$$

Therefore

$$P(m/p) = \binom{n}{m} \exp\{m \ln p + (n-m)\ln(1-p)\},$$

which we can write in the form

$$P(m/p) = \binom{n}{m} \exp\left\{\frac{m}{n}(n \ln p - n \ln(1-p)) + n \ln(1-p)\right\},$$

It is easy to see that

$$h(m) = \binom{n}{m}, \quad A(p) = n \ln p - n \ln(1-p), \quad B(p) = n \ln(1-p)).$$

Therefore $\frac{m}{n}$ is the effective estimate for unknown probability p.

Example 16.3. Estimating the mean of a Uniform distribution. Suppose $X_1, ..., X_n$ constitute a sample from a uniform distribution on $(0, \theta)$, where θ is unknown. Their joint density is thus

$$f(x_1, ..., x_n|\theta) = \begin{cases} \frac{1}{\theta^n}, & 0 < x_i < \theta, \quad i = 1, ..., n \\ 0 & \text{otherwise.} \end{cases}$$

This density is maximized by choosing θ as small as possible. Since θ must be at least as large as all of the observed values x_i, it follows that the smallest possible choice of θ is equal to $\max(x_1, x_2, ..., x_n)$. Hence, the maximum likelihood estimator of θ is

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

It easily follows from the foregoing that the maximum likelihood estimator of $\theta/2$, the mean of the distribution, is $\max(X_1, X_2, ..., X_n)/2$.