V ASYMPTOTICALLY MOST POWERFUL TESTS AND ASYMPTOTICALLY SHORTEST CONFIDENCE INTERVALS[9]

As we have seen, if a uniformly most powerful (unbiased) test and a shortest (unbiased) confidence interval exist, they provide a satisfactory solution of the problem of testing a hypothesis and the problem of interval estimation. Unfortunately, they exist only in a restricted class of cases. As substitutes for them the use of a critical region of type A and a short confidence interval, respectively, have been proposed. The appropriateness of the region of type A seems somewhat doubtful, since we are more interested in the behavior of the power function at values of θ far from the value θ_0 to be tested than at values of θ near to θ_0. Similar objections can be raised to the use of a short confidence interval. Recent investigations show, however, that the situation is much more favorable than appears at first glance. It is shown that the difficulties arising because of the non-existence of uniformly most powerful unbiased tests and shortest unbiased confidence intervals gradually disappear with increasing size of the sample, since so-called asymptotically most powerful unbiased tests and asymptotically shortest unbiased confidence intervals practically always exist.

We shall assume that the observations $x_1,\ldots,x_n$ are n independent observations on the same random variable X whose distribution function involves a single unknown parameter θ. We shall also assume that X has a probability density function,

---

9) See references 17-20

say $f(x,\theta)$. Since in our discussions the number of observations n will not be kept constant, we shall indicate the dimension of the sample space by proper subscripts. For instance, a critical region in the n-dimensional sample space will be denoted by a capital letter with the subscript n. A point of the n-dimensional sample space will be denoted by E_n, and a confidence interval based on n observations by $\delta_n(E_n)$.

For any region U_n denote by $G(U_n)$ the greatest lower bound of $P(U_n|\theta)$. For any pair of regions U_n and T_n denote by $L(U_n,T_n)$ the least upper bound of

$$P\left[U_n(\theta) - P(T_n|\theta)\right].$$

A sequence $\{W_n\}$ ($n=1,\ldots,$ad inf.) of regions is said to be *an asymptotically most powerful test of the hypothesis* $\theta = \theta_0$ *on the level of significance* α if $P(W|\theta_0) = \alpha$ and if for any sequence $\{Z_n\}$ of regions for which $P(Z_n|\theta_0) = \alpha$,

$$\lim_{n\to\infty} \sup L(Z_n,W_n) = 0 \quad \text{holds.}$$

A sequence $\{W_n\}$ ($n=1,\ldots,$ad inf.) of regions is said to be *an asymptotically most powerful unbiased test of the hypothesis* $\theta = \theta_0$ *on the level of significance* α if $P(W_n|\theta_0) = \lim_n G(W_n) = \alpha$ and if for any sequence $\{Z_n\}$ of regions for which $P(Z_n|\theta_0) = \lim G(Z_n) = \alpha$ the inequality $\lim \sup L(Z_n,W_n) \leq 0$ holds.

Let $P_n(\theta,\alpha)$ be defined by

$$P_n(\theta,\alpha) = \text{l.u.b. } P(Z_n|\theta)$$

with respect to all regions Z_n for which $P(Z_n|\theta_0) = \alpha$. We will call $P_n(\theta,\alpha)$ the envelope function corresponding to the level of significance α. Similarly let $P_n^*(\theta,\alpha)$ be the least upper bound of $P(Z_n|\theta)$ with respect to all unbiased critical regions Z_n which have the size α. We will call $P_n^*(\theta,\alpha)$ the unbiased envelope function corresponding to the level of significance α.

The two previously given definitions are equivalent to the following two:

A sequence $\{W_n\}$ of regions is said to be an asymptotically most powerful test of the hypothesis $\theta = \theta_0$ on the level of significance α if $P(W_n | \theta_0) = \alpha$ and

$$\lim_{n=\infty} \left\{ P_n(\theta,\alpha) - P(W_n | \theta) \right\} = 0$$

uniformly in θ.

A sequence $\{W_n\}$ of regions is said to be an asymptotically most powerful unbiased test of the hypothesis $\theta = \theta_0$ on the level of significance α if $P(W_n | \theta_0) = \alpha$ and

$$\lim_{n=\infty} \left\{ P_n^*(\theta,\alpha) - P(W_n | \theta) \right\} = 0$$

uniformly in θ.

Let $\hat{\theta}_n(x_1,\ldots,x_n)$ be the maximum likelihood estimate of θ in the n-dimensional sample space. That is to say, $\hat{\theta}_n$ denotes the value of θ for which the product $\prod_{\alpha=1}^{n} f(x_\alpha,\theta)$ becomes a maximum. Let W_n' be the region defined by the inequality $\sqrt{n}(\hat{\theta}_n = \theta_0) \geq c_n'$, W_n'' defined by the inequality $\sqrt{n}(\hat{\theta}_n - \theta_0) \leq c_n''$ and let W_n be defined by the inequality $|\sqrt{n}(\hat{\theta}_n - \theta_0)| \geq d_n$. The constants d_n, c_n', c_n'' are chosen in such a way that

$$P(W_n' | \theta_0) = P(W_n'' | \theta_0) = P(W_n | \theta_0) = \alpha.$$

It has been shown that under certain restrictions on the probability density $f(x,\theta)$ the sequence $\{W_n'\}$ is an asymptotically most powerful test of the hypothesis $\theta = \theta_0$ if θ takes only values $\geq \theta_0$. Similarly $\{W_n''\}$ is an asymptotically most powerful test if θ takes only values $\leq \theta_0$. Finally $\{W_n\}$ is an asymptotically most powerful unbiased test if θ can take any real value.

There are also other asymptotically most powerful tests. Let W_n' be the region defined by the inequality

$$\frac{1}{\sqrt{n}} \sum_{\alpha=1}^{n} \frac{\partial}{\partial \theta} \log f(x_\alpha, \theta_o) \geq c_n' ,$$

W_n'' defined by the inequality

$$\frac{1}{\sqrt{n}} \sum_{\alpha} \frac{\partial}{\partial \theta} \log f(x_\alpha, \theta_o) \leq c_n'' ,$$

and W_n defined by the inequality

$$\left| \frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} \sum_{\alpha} \log f(x_\alpha, \theta_o) \right| \geq c_n$$

where the constants c_n, c_n' and c_n'' are chosen in such a way that

$$P(W_n' | \theta_o) = P(W_n'' | \theta_o) = P(W_n | \theta_o) = \alpha.$$

Then $\{W_n'\}$ is an asymptotically most powerful test of the hypothesis $\theta = \theta_o$ if θ takes only values $\geq \theta_o$. Similarly, $\{W_n''\}$ is an asymptotically most powerful test if θ takes only values $\leq \theta_o$. Finally $\{W_n\}$ is an asymptotically most powerful unbiased test if θ can take any real value.

The sequence $\{A_n(\theta_o)\}$ is an asymptotically most powerful unbiased test of the hypothesis $\theta = \theta_o$, where $A_n(\theta_o)$ denotes the critical region of type A for testing the hypothesis $\theta = \theta_o$.

Since there are many asymptotically most powerful tests, the question arises whether they are all equally good or whether one can be preferred to another. It is clear that if $\{W_n\}$ and $\{W_n'\}$ are two asymptotically most powerful unbiased tests, then for sufficiently large n they are equally good. In fact, for sufficiently large n both power functions $P(W_n | \theta)$ and

$P(W_n'|\theta)$ are in a small neighborhood of $P_n(\theta,\alpha)$ $\left[P_n^*(\theta,\alpha)\right]$. However, they may behave differently in the sense that with increasing n one power function, say $P(W_n|\theta)$ approaches the envelope function faster than $P(W_n'|\theta)$ does. In such a case it seems preferable to use W_n, especially if the sample is only moderately large. If the sample is so large that both power functions are in a small neighborhood of the envelope function, then it is immaterial whether we use W_n or W_n'.

These considerations lead to the idea that it is preferable to use that asymptotically most powerful (unbiased) test $\{W_n\}$ for which the approach of $P(W_n|\theta)$ to the envelope function is, in a certain sense, fastest.

A region W_n is called a most stringent test of size α for testing the hypothesis $\theta = \theta_0$ if $P(W_n|\theta_0) = \alpha$ and

$$\underset{\theta}{\text{l.u.b.}}\left[P_n(\theta,\alpha)-P(W_n|\theta)\right] \quad \underset{\theta}{\text{l.u.b.}}\left[P_n(\theta,\alpha)-P(Z_n|\theta)\right]$$

for all Z_n for which $P(Z_n|\theta_0) = \alpha$. The abbreviation $\underset{\theta}{\text{l.u.b.}}$ means "least upper bound with respect to θ."

If W_n is for each n a most stringent test, its power function will approach the envelope function, in a certain sense, faster than any other power function. It seems, therefore, desirable to use a most stringent test. A region of type A is not exactly a most stringent test, but probably it is quite near to it (this question has yet to be investigated), and this would provide a very good justification for the use of a type A region. The mathematical difficulties in finding explicitly a most stringent test are considerable.

Let $\delta_n(E_n) = [\underline{\theta}_n(E_n), \bar{\theta}_n(E_n)]$ be an interval function and denote by $P[\delta_n(E_n) \, C\theta' | \theta'']$ the probability that $\delta_n(E_n)$ will cover θ' under the assumption that θ'' is the true value of the parameter.

A sequence of interval functions $\{\delta_n(E_n)\}$ (n=1,2,...,ad inf) is called <u>an asymptotically shortest confidence interval of θ</u> if the following two conditions are fulfilled:

(a) $P[\delta_n(E_n) \quad C\theta|\theta] = \alpha$ for all values of θ

(b) For any sequence of interval functions $\{\delta_n'(E_n)\}$ (n=1,2,..., ad inf.) which satisfies (a), the least upper bound of

$$P[\delta_n(E_n) \, C\theta' | \theta''] - P[\delta_n'(E_n) \, C\theta' | \theta'']$$

with respect to θ' and θ'' converges to zero with $n \to \infty$.

A sequence of interval functions $\{\delta_n(E_n)\}$ (n=1,2,...,ad inf) is called <u>an asymptotically shortest unbiased confidence interval of θ</u> if the following three conditions are fulfilled:

(a) $P[\delta_n(E_n) \, C\theta|\theta] = \alpha$ for all values of θ

(b) The least upper bound of $P[\delta_n(E_n) \, C\theta' | \theta'']$ with respect to θ' and θ'' converges to α with $n \to \infty$

(c) For any sequence of interval functions $\{\delta_n'(E_n)\}$ which satisfies the conditions (a) and (b), the least upper bound of

$$P[\delta_n(E_n) \, C\theta' | \theta''] - P[\delta_n'(E_n) \, C\theta' | \theta'']$$

with respect to θ' and θ'', converges to zero with $n \to \infty$.

Let $C_n(\theta)$ be a positive function of θ such that the probability that $\left| \frac{1}{\sqrt{n}} \sum_\beta \frac{\partial}{\partial \theta} \log f(x_\beta, \theta) \right| \leq C_n(\theta)$ is equal to a constant α under the assumption that θ is the true value of the parameter. Denote by $\underline{\theta}(E_n)$ the root in θ of the equation

$$\frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} \sum_{\beta} \log f(x_\beta, \theta) = C_n(\theta)$$ and by $\overline{\theta}(E_n)$ the root of

$$\frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} \sum_{\beta} \log f(x_\beta, \theta) = -C_n(\theta).$$ It has been shown that under some restrictions on $f(x,\theta)$ the interval $\delta(E_n) = [\underline{\theta}(E_n), \overline{\theta}(E_n)]$ is an asymptotically shortest unbiased confidence interval of θ corresponding to the confidence coefficient α. This confidence interval is identical with that given by Wilks[10]. The definition of a shortest confidence interval underlying Wilks' investigations is somewhat different from that of Neyman's, which has been used here. According to Wilks, a confidence interval $\delta(E)$ is called shortest in the average if the expectation of the length of $\delta(E)$ is a minimum. The main result obtained by Wilks can be formulated as follows: The confidence interval in question is asymptotically shortest in the average compared with all confidence intervals the endpoints of which are roots of an equation of the following type:

$$\sum_{\beta} h(x_\beta, \theta) = \pm C_n(\theta).$$

In the present investigation such a restriction is not made. The confidence interval in consideration is shown to be asymptotically shortest compared with any unbiased confidence interval.

Now let $C_n(\theta)$ be a positive function of θ such that the probability that $|\hat{\theta}_n - \theta| \leq C_n(\theta)$ is equal to a constant α under

10) See reference 22

the assumption that θ is the true value of the parameter. Denote by $\underline{\theta}(E_n)$ the root in θ of the equation $\hat{\theta}_n - \theta = C_n(\theta)$ and by $\overline{\theta}(E_n)$ the root of $\hat{\theta}_n - \theta = -C_n(\theta)$. Consider the interval $\delta(E_n) = \left[\underline{\theta}(E_n), \overline{\theta}(E_n)\right]$. Under some restrictions on the density $f(x,\theta)$, it can be shown that $\delta(E_n)$ is an asymptotically shortest unbiased confidence interval.

This is a much stronger property of the maximum likelihood estimate than its efficiency and gives a justification of the use of the maximum likelihood estimate also in the light of Neyman's theory of estimation.