

ON BEST ASYMPTOTIC CONFIDENCE INTERVALS FOR PARAMETERS OF STOCHASTIC PROCESSES

BY C. C. HEYDE

Australian National University

This paper is concerned with the size of confidence intervals for parameters of stochastic processes based on limit laws with two competing normalizations, one producing asymptotic normality and the other asymptotic mixed normality. It is shown that, in a certain sense, the interval based on asymptotic normality is preferable on average. Applications to estimation of parameters in nonergodic stochastic processes and to estimation of steady-state parameters in a simulation are given to illustrate the theory.

1. Introduction and results. Much asymptotic inference for stochastic processes is based on use of some obvious consistent estimator but involves a choice between competing normalizations, one being of constants and the other of random variables. In this paper we study the common situation where either asymptotic normality or asymptotic mixed normality is achievable through suitable normalization. It is shown that there is a certain sense in which, on average, confidence intervals based on the asymptotic normality result are preferable to those based on asymptotic mixed normality.

THEOREM. *Let $\{\hat{\theta}_n\}$ be a sequence of estimators which is consistent for θ and $\{c_n\}, \{d_n\}$ be norming sequences, possibly random, such that*

$$(1) \quad c_n(\hat{\theta}_n - \theta) \rightarrow_d W,$$

$$(2) \quad d_n(\hat{\theta}_n - \theta) \rightarrow_d \eta^{-1}W$$

as $n \rightarrow \infty$, where W is standard normal and $\eta > 0$ a.s. is random, independent of W and such that

$$c_n d_n^{-1} \rightarrow_d \eta$$

as $n \rightarrow \infty$. Suppose that $L_n^{(i)}(\delta)$, $i = 1, 2$, is the minimum length of an $100(1 - \delta)\%$ confidence interval for θ based on an exact, approximate or assumed distribution of the pivot in (1) and (2), respectively, for which the specified convergence result holds. Then

$$\liminf_{n \rightarrow \infty} E(L_n^{(2)}(\delta)/L_n^{(1)}(\delta)) \geq 1.$$

Received January 1991; revised July 1991.

AMS 1980 subject classifications. Primary 62F11; secondary 62F25, 62M09.

Key words and phrases. Best confidence intervals, normalization, asymptotic normality, asymptotic mixed normality, nonergodic models.

REMARK 1. The theorem gives a sense in which, on average, confidence intervals based on the pivot in (1) are to be preferred to those based on the pivot in (2) whether or not a random norming is required.

REMARK 2. A principal application of the theorem is in the context of inference for the nonergodic models that are common in stochastic process estimation [e.g., Basawa and Scott (1983), Hall and Heyde (1980), Chapter 6]. In this context we have $\{c_n\}$ as random and $\{d_n\}$ as constants so that $L_n^{(1)}(\delta)$ is a random variable and $L_n^{(2)}(\delta)$ is a constant.

One of the most important special cases is that of locally asymptotic mixed normal (LAMN) families. This is the situation in which the log-likelihood ratio has asymptotically a mixed normal distribution and results of type (2) hold with $\{d_n\}$ as constants and (1) with $\{c_n\}$ as random variables, while $Ec_n = d_n$ for each n . For LAMN families one has, under modest regularity conditions, a striking result known as Hajék's convolution theorem [e.g., Theorem 2, page 47, of Basawa and Scott (1983)] to the effect that if $\{T_n\}$ is any other sequence of consistent estimators of θ such that

$$(3) \quad d_n(T_n - \theta) \rightarrow_d U$$

for some nondegenerate U , then

$$U =_d V + \eta^{-1}W,$$

where V is independent of η and W .

It is readily shown that confidence intervals based on the pivot in (3) are wider than those for the corresponding result (2). That is, using (1) is better on average than (2), which is better than (3). If Φ is the standard normal distribution function, we have for any real α, β, α with $\beta > \alpha$,

$$(4) \quad \Phi((\beta - \alpha)y) - \Phi((\alpha - \alpha)y) \leq \Phi(\frac{1}{2}(\beta - \alpha)y) - \Phi(-\frac{1}{2}(\beta - \alpha)y)$$

and hence, integrating with respect to $dP(\eta \leq y)$,

$$(5) \quad P(\alpha < \eta^{-1}W + a < \beta) \leq P(-\frac{1}{2}(\beta - \alpha) < \eta^{-1}W < \frac{1}{2}(\beta - \alpha)),$$

so that

$$\begin{aligned} P(\alpha < \eta^{-1}W + V < \beta) &= \int_{-\infty}^{\infty} P(\alpha < \eta^{-1}W + v < \beta) dP(V \leq v) \\ &\leq P(-\frac{1}{2}(\beta - \alpha) < \eta^{-1}W < \frac{1}{2}(\beta - \alpha)). \end{aligned}$$

REMARK 3. A particular case of the result of the theorem has been obtained by Glynn and Iglehart (1990) who studied the problem of finding minimum size asymptotic confidence intervals for steady state parameters of the simulation output process from a single simulation run. They contrasted the approach of consistently estimating the variance constant in the relevant central limit theorem with the standardized time series approach which avoids estimation of the variance in a manner reminiscent of the t -statistic and suggested that the former approach is preferable on average.

2. Proof of theorem. Suppose that Φ_n and Ψ_n are exact, approximate or assumed distribution functions of the pivots $c_n(\hat{\theta}_n - \theta)$ and $d_n(\hat{\theta}_n - \theta)$, respectively. We are given the complete convergence results

$$\Phi_n \rightarrow_c \Phi, \quad \Psi_n \rightarrow_c \Psi$$

as $n \rightarrow \infty$, where

$$\Phi(x) = P(W \leq x) \quad \text{and} \quad \Psi(x) = P(\eta^{-1}W \leq x) = \int_0^\infty \Phi(xy) dG(y)$$

with $G(x) = P(\eta \leq x)$.

Now let a_n, b_n be any numbers for which

$$\Phi_n(b_n) - \Phi_n(a_n) \geq 1 - \delta.$$

This gives a confidence interval for θ of length

$$l_n^{(1)}(\delta) = (b_n - a_n)/c_n.$$

By passing to a subsequence, we may suppose that $a_n \rightarrow a$ and $b_n \rightarrow b$, where $-\infty \leq a \leq b \leq \infty$. Then, by complete convergence,

$$\Phi_n(a_n) - \Phi_n(b_n) \rightarrow \Phi(b) - \Phi(a),$$

and, in view of (4), $b - a \geq 2z_\delta$, where $2\Phi(z_\delta) = 2 - \delta$. Therefore, noting that $c_n l_n^{(1)}(\delta)$ is not random, we have

$$(6) \quad \liminf_{n \rightarrow \infty} c_n L_n^{(1)}(\delta) \geq 2z_\delta.$$

Next, let z_n be the smallest z for which

$$\Phi_n(z) - \Phi_n(-z) \geq 1 - \delta.$$

As above, we may suppose that $z_n \rightarrow z_\infty$ and, because Φ is continuous,

$$\Phi_n(z_n) - \Phi_n(-z_n) = 1 - \delta + o(1)$$

as $n \rightarrow \infty$. By complete convergence and using the result $\Phi(-x) = 1 - \Phi(x)$, $x > 0$, it follows that $2\Phi(z_\infty) = 2 - \delta$ and hence that $z_\infty = z_\delta$. Then, $L_n^{(1)}(\delta) \leq 2z_n/c_n$, so that

$$(7) \quad \liminf_{n \rightarrow \infty} c_n L_n^{(1)}(\delta) \leq 2z_\delta,$$

and (6) and (7) imply

$$(8) \quad \lim_{n \rightarrow \infty} c_n L_n^{(1)}(\delta) = 2z_\delta.$$

Similar reasoning to that which led to (8) also applies to show that

$$(9) \quad \lim_{n \rightarrow \infty} d_n L_n^{(2)}(\delta) = 2z_\delta,$$

where $2\Psi(z_\delta) = 2 - \delta$. We merely replace (4) by (5) to show that the symmetric confidence interval is best, while $\Psi(-x) = 1 - \Psi(x)$, $x > 0$, is evident from the definition and the corresponding result for Φ .

From (6) and (9), it then follows that, as $n \rightarrow \infty$,

$$d_n L_n^{(2)}(\delta) / c_n L_n^{(1)}(\delta) \rightarrow \zeta_\delta / z_\delta.$$

Since $c_n / d_n \rightarrow_d \eta$, Slutsky's theorem gives

$$L_n^{(2)}(\delta) / L_n^{(1)}(\delta) \rightarrow_d \eta \zeta_\delta / z_\delta$$

and then, using Billingsley [(1968), Theorem 5.3, page 32],

$$\liminf_{n \rightarrow \infty} E(L_n^{(2)}(\delta) / L_n^{(1)}(\delta)) \geq (E\eta) \zeta_\delta / z_\delta.$$

Thus, in order to complete the proof, it remains to show that

$$(10) \quad (E\eta) \zeta_\delta \geq z_\delta.$$

Now note that $\zeta_\delta = \zeta(\eta; \delta)$ solves the equation

$$\Psi(\zeta(\eta; \delta)) = P(\eta^{-1}W \leq \zeta(\eta; \delta)) = 1 - \delta/2.$$

Then, taking $b > 0$, we have

$$1 - \frac{\delta}{2} = P(W \leq \zeta(\eta; \delta)\eta) = P\left(W \leq \frac{1}{b}\zeta(\eta; \delta)b\eta\right) = P(W \leq \zeta(b\eta; \delta)b\eta)$$

so that continuity and strict monotonicity of Ψ imply that

$$\zeta(b\eta; \delta) = \frac{1}{b}\zeta(\eta; \delta)$$

and hence, if

$$\psi(\eta) = (E\eta)\zeta(\eta; \delta),$$

we have

$$(11) \quad \psi(b\eta) = \psi(\eta).$$

Now we see from (10) that it is required to show

$$\psi(\eta) \geq \Phi^{-1}(1 - \delta/2)$$

and using (11) we may, without loss of generality, scale η so that

$$(12) \quad \zeta(\eta; \delta) = 1.$$

But (12) implies

$$\Psi(1) = 1 - \delta/2,$$

or equivalently,

$$\int_0^\infty \Phi(y)G(y) = E\Phi(\eta) = 1 - \delta/2.$$

Thus, we have to show that

$$\psi(\eta) = E\eta \geq \Phi^{-1}(1 - \delta/2)$$

subject to

$$E\Phi(\eta) = 1 - \delta/2,$$

and since Φ is monotone, this holds if

$$(13) \quad \Phi(E\eta) \geq E\Phi(\eta).$$

However, since $1 - \Phi(x)$ is convex for $x > 0$, as is easily checked by differentiation, we have from Jensen's inequality that

$$1 - \Phi(E\eta) \leq E(1 - \Phi(\eta))$$

and (13) follows. This completes the proof.

FINAL REMARKS. It is worth noting that (10), which relates the average confidence intervals when (1) and (2) hold exactly rather than as limits, is a paraphrase of the fact that, on average, extraneous randomization does not improve confidence intervals for a normal mean. That is, if X is distributed as $N(\theta, 1)$, the usual symmetric confidence interval for θ based on $X - \theta$ is not improved, on average, if one instead uses the pivot $(x - \theta)/\eta$, where η is independent of θ . In particular, when \bar{X}_n, s_n^2 are the mean and variance of a random sample from the $N(\theta, \sigma^2)$ distribution where σ is known, using the t -intervals for θ based on $(\bar{X}_n - \theta)/s_n$ produces confidence intervals which are longer on average than z -intervals.

Acknowledgment. The perceptive comments of an Associate Editor which purged the theorem of a number of unnecessary restrictions are gratefully acknowledged.

REFERENCES

- BASAWA, I. V. and SCOTT, D. J. (1983). *Asymptotic Optimal Inference for Non-ergodic Models. Lecture Notes in Statist.* **17**. Springer, New York.
- BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- GLYNN, P. and IGLEHART, D. L. (1990). Simultaneous output analysis using standardized time series. *Math. Oper. Res.* **15** 1-16.
- HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic, New York.

STATISTICS RESEARCH SECTION
INSTITUTE OF ADVANCED STUDIES
AUSTRALIAN NATIONAL UNIVERSITY
GPO BOX 4, CANBERRA
ACT 2601
AUSTRALIA