OMNIBUS CONFIDENCE INTERVALS

Michael Sherman and Edward Carlstein*
Texas A&M University and University of North Carolina at Chapel Hill

Abstract:

Confidence interval construction for an unknown parameter typically requires knowledge of the convergence rate of a point-estimator targeting the parameter and knowledge of the asymptotic distribution of the [standardized] point-estimator. A general subsampling method is presented here for obtaining confidence intervals which requires specification of neither the estimator's asymptotic distribution nor its convergence rate; the unknown asymptotic distribution may be non-Normal, and the unknown convergence rate may differ from the familiar $n^{1/2}$. Serial dependence is allowed in the observed data sequence, and the dependence mechanism can be unknown. Under mild conditions, the intervals asymptotically obtain the nominal coverage level and interval-width shrinks to zero as sample size increases. Finite-sample behavior of the omnibus confidence interval is studied by example and by simulation.

Keywords: Dependence, Heavy-Tails, Nonparametric, Resampling, Subsampling, Time-Series

---

* Michael Sherman is Assistant Professor in the Department of Statistics, Texas A&M University, College Station, TX 77843, and Edward Carlstein is Professor in the Department of Statistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599. We are grateful to Patrice Bertail, Dimitris N. Politis, and Joseph P. Romano for sharing with us an early version of their paper.

1. INTRODUCTION

Consider the problem of constructing a confidence interval for an unknown real-valued "target parameter" θ, using n observations from a stationary random sequence. From the data $\{X_1, X_2, ..., X_n\}$ a statistic $s_n := s_n(X_1, X_2, ..., X_n)$ is computed as a point-estimator of θ. In order to parlay this point-estimator into a confidence interval for θ, the user typically needs knowledge about the sampling distribution of s_n, or, as a practical approximation, knowledge about the asymptotic distribution (F) of a standardized transform of s_n, say, $a_n(s_n - \theta)$, where $a_n > 0$. The user's level of knowledge [or assumptions] about F and a_n will naturally influence the choice of method for confidence interval construction.

In the ideal case, F and a_n are both completely known. Then the usual pivot yields the nominal two-sided equi-tailed $\beta(100)\%$ confidence interval

$$I_n^\beta(F, a_n) := [s_n - F^{-1}((1+\beta)/2)/a_n,\ s_n - F^{-1}((1-\beta)/2)/a_n],$$

which has correct asymptotic coverage probability $\beta \in (0,1)$. In practice, however, this idealized situation rarely occurs. Even in the familiar scenario where F is Normal and a_n is of the order $n^{1/2}$, there is still typically an unknown scale parameter σ which must be estimated. More serious obstacles also frequently arise:

(i) The statistic s_n may be complicated (e.g., a robustified or adaptively defined statistic, like Switzer's adaptive trimmed-mean [see Efron (1982), Example 5.2]), so that a theoretical derivation of F is analytically intractable.

(ii) The observations might be serially dependent, so their joint probability structure must be accounted for in deriving F; this in turn may require knowledge or assumptions about the underlying serial dependence mechanism, and estimation of its concomitant parameters (e.g., in an assumed AR(p) model, the noise distribution as well as the p autoregressive coefficients all could be relevant in deriving F).

When difficulties such as these render F unavailable, it may be possible to construct an "estimated" confidence interval via the jackknife histogram (Wu (1990)), the bootstrap (Efron (1979)), or the subsampling method of Politis and Romano (1994) [referenced as PR in the sequel]. The jackknife histogram is justified only when the X_i's are i.i.d. and the statistic is asymptotically Normal with $a_n = n^{1/2}$. The bootstrap has been justified in a wide variety of situations, including asymptotically Normal (Bickel and Freedman (1981), Singh (1981)) and non-Normal statistics (Athreya (1987), Bretagnolle (1983), Swanepoel (1986)), and allowing for serially dependent X_i's (Basawa et al. (1989), Bose (1988), Künsch (1989), Rajarshi (1990)); but the construction and theoretical justification of bootstrap confidence intervals generally requires that the user have knowledge about the standardizing sequence a_n for the specific statistic s_n at hand. PR's method again yields a broadly applicable [estimated] confidence interval $I_n^{\beta}(\hat{F}, a_n)$ – provided that a_n is known as a function of n.

1.1 What Happens If The Standardizing Sequence a_n Is Unknown?

Specifically, a_n may involve unknown parameters, or, a_n may not even be known in functional form. Consider the following two classes of examples, which will be referred to repeatedly in the sequel.

Example 1: When s_n is the usual sample mean $(\overline{X}_n)$ of i.i.d. observations, then F is necessarily an α-stable distribution $(0 < \alpha \leq 2)$ and $a_n = n^{(\alpha-1)/\alpha} h(n)$, where $h(\cdot)$ is a "slowly varying" function (i.e., $h(cx)/h(x) \to 1$ as $x \to \infty$ for each fixed $c > 0$). The familiar case of $\alpha = 2$, F=Normal, $h(n) = constant$, is just one out of a wide spectrum of possibilities including skewed F and $h(n) \to \infty$ or $h(n) \to 0$ [see Ibragimov and Linnik (1971), Chapter 2, for a complete analysis]. Inference for the center or location (θ) when $\alpha < 2$ is closely tied to the classical study of robustness in the presence of heavy-tailed data; in this context, it has long been recognized that, in applications, a_n is likely to be unknown (e.g., Fama and Roll (1968)). There are estimates of α available (e.g., Hall (1982)) for use in this particular situation, although direct estimation of unknown $h(n)$ seems not to be dealt with in the literature. Moreover, none of the other existing methods intended for omnibus confidence interval construction (e.g., bootstrap) can handle, in an automatic way, this rather fundamental example with α and $h(n)$ unknown.

Example 2: Suppose that s_n is a U-statistic computed from i.i.d. observations [see Serfling (1980), Chapter 5]. When the U-statistic is "non-degenerate", then $a_n = n^{1/2}$ and F is a Normal distribution. But when the U-statistic is "degenerate", then $a_n = n$ and F is the distribution of a weighted sum of χ^2 random variables. The degenerate case is not an obscure theoretical anomaly: Several common goodness-of-fit statistics are degenerate U-statistics; and, moreover, a single U-statistic (s_n) can behave as either degenerate or non-degenerate according to quite subtle changes in the underlying distribution of the data [see Section 4.2].

More generally, unknown a_n will typically arise in conjunction with obstacle (i) [discussed above]: When theoretical derivation of F is intractable, then, realistically, the proper standardization sequence $\{a_n\}$ will also often be unavailable to the user. Note that determination of the proper $\{a_n\}$ is an inherently theoretical problem requiring situation-specific analysis of the particular s_n at hand. As an illustration of this dilemma, consider PR's Example 2.1.1: The example involves a somewhat complicated statistic s_n and parameter θ, but no progress can be made towards constructing their subsampling-based confidence interval without invoking an earlier theoretical derivation of $a_n = n^{1/3}$.

Bertail, Politis, and Romano (1995) [referenced as BPR in the sequel] have recognized that this issue of unknown a_n is crucial to the applicability of confidence interval construction methods; they argue convincingly for the need to develop methods which avoid explicit knowledge of a_n. In the special case where it is known that $a_n = n^r$ $(r > 0)$, BPR provide consistent estimators $(\hat{r})$ of the unknown r, which are then "plugged-in" yielding an interval of the form $I_n^\beta(\widehat{\hat{F}}, n^{\hat{r}})$. Although their interval achieves asymptotically the correct coverage probability, it does leave several issues unresolved:

(I) For precisely the same reasons that motivate BPR's method, it is *a fortiori* desirable to have omnibus confidence intervals applicable in the general case where a_n may not be n^r. In particular, to fully treat even the basic statistic $\overline{X}_n$ [Example 1, above], one must allow $h(n)$ to be a nonconstant slowly varying function. BPR dismiss as infeasible the task of estimating the slowly varying function − perhaps foreshadowing the principle that a truly omnibus interval should bypass any attempt at direct estimation of general unknown a_n.

(II) Even when the assumption $a_n = n^r$ holds, plug-in estimation of r in $I_n^\beta(\widehat{\hat{F}}, n^{\hat{r}})$ introduces an additional source of sampling variability into the interval. Also, implementation of BPR's rate estimate $\hat{r}$ requires the choice of several auxiliary "tuning parameters" (their I, J, $\{b_{i,n}\}, \{t_j\}$). These practical difficulties can again be avoided if direct estimation of a_n is not attempted.

1.2 Omnibus Confidence Intervals

An "omnibus" procedure should apply to a general statistic in a general setting, so that each new scenario (e.g., a new statistic s_n, a new standardizing sequence a_n, or a new serial dependence mechanism) does not require the development of a new methodology. The potential utility of such procedures motivated the jackknife and the bootstrap; Efron (1982, p.28) finds the "charm" of these procedures in that "they can be applied to complicated situations where parametric modeling and/or theoretical analysis is hopeless".

It is, however, essential to distinguish between *ad hoc* application versus rigorous justification of proposed procedures. This distinction is brought sharply into focus by Young's (1994) review paper on the bootstrap, where it is stated that "establishing validity of bootstrap in a particular setting may be a highly nontrivial exercise" (p.386), and "in many settings there is still much theoretical analysis of bootstrap required before we can be confident of its value" (p.391). "The danger", according to Young (p.385), "is that the practitioner may well be attracted to use of the bootstrap...in precisely those circumstances where the bootstrap is most likely to fail and where least is known". Indeed, situations involving non-standard [or unknown] a_n are prominent illustrations of how attempts at "automatic" implementation of the naive bootstrap can fail (see, e.g., Young (1994) and BPR). Although the bootstrap can achieve second-order accuracy in certain cases, the pursuit of omnibus confidence intervals gives primacy to broad validity [of first-order properties] with minimal user knowledge/assumptions. This is the philosophy of BPR, who "opt for a minimalist approach"; "the spirit" of their work "is to obtain valid results under the weakest assumptions possible". The present paper continues in the same spirit.

The omnibus confidence interval presented in Section 2 does achieve first-order accuracy (see Theorem) and does so *without imposing assumptions on the specific functional form of* a_n and *without attempting to explicitly estimate the unknown* a_n. Because the unknown a_n is allowed to be completely general, this new confidence interval is more broadly applicable than previously existing methods [cf. (I) above]. Also, there is no need to estimate any unknown nuisance parameters (like r) in a_n, so the new method is simple to implement and avoids the practical difficulties discussed in (II) (above).

1.3 The Role of Serial Dependence

The presence of any serial dependence in the data $\{X_1, X_2, ..., X_n\}$ will typically exacerbate the analytical difficulty of determining F and a_n [see obstacle (ii) above]. In this context "the rewards may be higher, but the problems are trickier...", as noted by Young (1994, p.389). The qualitative behavior encountered in Examples 1 and 2 (above) persists when the data are serially dependent [see, e.g., Ibragimov and Linnik (1971), Chapter 18, and Carlstein (1988)]. To handle possible serial dependence in an omnibus spirit, the proposed confidence interval employs "subseries" of data, which automatically retain the correct dependence structure without user knowledge of the underlying dependence mechanism (e.g., ARMA(p, q)). The only assumption regarding the strength of dependence is a mild model-free "mixing" condition.

1.4 The Cost

Any method which achieves such broad applicability (i.e., general statistics s_n; unknown a_n; serially dependent data; minimal regularity conditions; simple implementation) must come at some cost. The cost here is that the proposed omnibus confidence interval does not have optimal width, compared to what could be obtained using parametric or theoretical knowledge (e.g., $I_n^{\beta}(F, a_n)$). Asymptotically, the width of the proposed interval does shrink to zero (see Theorem).

1.5 Outline of the Article

Section 2 explicitly defines the omnibus confidence interval and presents its theoretical properties [asymptotic validity and width shrinkage]. A natural tradeoff between accuracy (i.e., actual coverage probability relative to nominal) and interval-width is explored in Section 3. Simulations (Section 4) illustrate the finite-sample performance of omnibus confidence intervals in two nontrivial situations: the sample mean of heavy-tailed observations, with nonconstant $h(n)$ in a_n (Section 4.1); and a simple statistic $s_n(\cdot)$ [computed on serially dependent data] which can behave either as approximately Normal or as drastically non-Normal (Section 4.2).

2. OMNIBUS CONFIDENCE INTERVALS: DEFINITIONS AND PROPERTIES

A "general statistic" is determined by a sequence of completely known functions $\{s_m(\cdot) : m \geq 1\}, s_m(\cdot) : (R^d)^m \rightarrow R^1$, where $d \geq 1$ is the dimension of X_i. In principle, a general statistic $s_m(\cdot)$ allows for an arbitrarily complicated function of the data.

Introduce the following notation for a "subseries" of l consecutive observations:

$$X_l^i := (X_{i+1}, X_{i+2}, ..., X_{i+l}),$$

so that the observed data-set is X_n^0 and the collection of all available subseries is $\{X_l^i : 0 \le i \le n-l\}$. The associated "replicates" of the statistic are denoted by $s_l^i := s_l(X_l^i)$ [in the notation of Section 1, $s_m \equiv s_m^0$]. By employing subseries replicates s_l^i, the correct serial dependence structure is automatically retained without any knowledge or assumptions about the underlying dependence mechanism. Effective use of subseries replicates has been made in other resampling methods, e.g., BPR, Carlstein (1986), and PR.

As an empirical analog of the unknown $P\{s_l - \theta \le y\}$, calculate the distribution:

$$G_{l,n}(y) := \sum_{i=0}^{n-l} \mathbf{1}\{s_l^i - s_n \le y\}/(n-l+1), \ y \in R.$$

The proposed omnibus confidence interval for θ is then simply:

$$I_l^\beta(G_{l,n}, 1) = [s_l - G_{l,n}^{-1}((1+\beta)/2), s_l - G_{l,n}^{-1}((1-\beta)/2)],$$

where the t^{th} quantile $(0 < t < 1)$ of any distribution function $G(\cdot)$ is formally defined as $G^{-1}(t) := \inf\{y : G(y) \ge t\}$. This interval is two-sided and equi-tailed with nominal coverage probability β, but other versions (e.g., non-equi-tailed, one-sided) can be similarly constructed using the regions $A_n(t)$ [see Appendix A.1].

The interval $I_l^\beta(G_{l,n}, 1)$ is directly computable from the available data, and does not require the user to specify the asymptotic distribution F, the standardization coefficient a_n, nor the serial dependence mechanism generating the data sequence. In order to establish the asymptotic properties of the interval, a few minimal regularity conditions (R.1 through R.3) will be introduced and discussed in the next subsection.

2.1 Notation and Conditions

In the asymptotic context $(n \to \infty)$, subseries length depends on sample size, i.e., $l = l(n)$. The strength of serial dependence in $\{X_i : -\infty < i < \infty\}$ is measured by the standard model-free "mixing coefficient":

$$\alpha(m) := \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}(..., X_{-1}, X_0), B \in \mathcal{F}(X_m, X_{m+1}, ...)\},$$

as defined by Rosenblatt (1956). Denote $y_0 := \sup\{y : F(y) = 0\}$ and $y_1 := \inf\{y : F(y) = 1\}$, as in BPR.

In order to establish the asymptotic properties of the omnibus confidence interval, consider the following minimal regularity conditions:

(R.1) $lim_{n\to\infty} P\{a_n(s_n - \theta) \leq y\} = F(y), \;\; y \in R,$ where $a_n \to \infty$ and $F(y)$ is continuous in y and strictly increasing on (y_0, y_1).

(R.2) $l \to \infty$, $l/n \to 0$, and $a_l/a_n \to 0$.

(R.3) $lim_{m\to\infty} \alpha(m) = 0$.

It is natural to have $a_n \to \infty$ in R.1 whenever s_n is a consistent estimator of θ. The existence of an asymptotic distribution $F(y)$ in R.1 is "almost a *sine qua non*" for the scenario presently being studied (BPR). The continuity and strict increase in R.1 are also needed by BPR to validate their method (which applies only in the special case $a_n = n^r$). When constructing empirical distributions (like $G_{l,n}$) from subseries replicates (s_l^i), it is standard to require R.2 and R.3 (see, e.g., PR and BPR). Increasing subseries length ($l \to \infty$) guarantees that the distributional behavior of replicates (s_l^i) approaches the asymptotic behavior (characterized by F). The condition $l/n \to 0$ provides a sufficiently large number of replicates for constructing $G_{l,n}$. Whenever $a_n = n^r h(n)$ [$r > 0, h(\cdot)$ slowly varying] and $l = \lfloor n^\gamma \rfloor$ $(0 < \gamma < 1)$, then $a_l/a_n \to 0$ is immediately satisfied; this broad class of a_n's includes the BPR set-up as a particular case, but also includes Example 1 (the sample mean) with nonconstant $h(n)$. When constructing the proposed interval in a situation where $a_n = n^r h(n)$, r and $h(\cdot)$ need not be known (cf., PR) nor do they need to be estimated (cf., BPR). Intuitively, the mixing condition R.3 says that replicates separated by a large time-lag behave approximately as if they were independent; note that R.3 is an especially mild mixing condition, as no rate of decay is imposed on $\alpha(\cdot)$.

2.2 Main Result

The following Theorem justifies the use of the proposed omnibus confidence intervals.

Theorem:

If R.1 *through* R.3 *hold, then as* $n \to \infty$:

$$P\{\theta \in I_l^\beta(G_{l,n}, 1)\} \to \beta \quad \text{and} \quad width[I_l^\beta(G_{l,n}, 1)] \xrightarrow{p} 0.$$

The proof [see Appendix A.1] leans heavily on the theory developed by PR and BPR.

3. THE EFFECT OF SUBSERIES LENGTH (l)

The proposed omnibus confidence interval $I_l^\beta(G_{l,n}, 1)$ has desirable coverage and width properties, as established in the above Theorem, for any subseries length l in the broad class allowed by R.2. But, for fixed finite sample size n, the particular choice of l will influence the coverage and width behavior of the interval. The inherent tradeoffs in this relationship are illustrated via exact calculations, in the following example.

3.1 Example 1, Revisited

Here we obtain exact computable expressions for coverage probability error and expected interval-width, for arbitrary sample size (n), subseries length (l), and nominal confidence coefficient (β), in the following situation: Let $\{X_i\}$ be i.i.d. with $P\{X_i = 1\} = 1/2 = P\{X_i = -1\}$, and consider the sample mean $s_m(X_1, X_2, ..., X_m) = \sum_{i=1}^m X_i/m$ as a point-estimator of $\theta = E\{X_i\} = 0$. For this example it will be more convenient to define

$$\tilde{G}_{l,n}(y) := \sum_{1 \le i_1 < i_2 \ldots < i_l \le n} \mathbf{1}\{\tilde{s}_l^i - s_n \le y\} / \binom{n}{l},$$

where $i := (i_1, i_2, ..., i_l)$ and $\tilde{s}_l^i := s_l(X_{i_1}, X_{i_2}, ..., X_{i_l})$, rather than our $G_{l,n}(y)$ [in the case of i.i.d. observations, $\tilde{G}_{l,n}(y)$ is a reasonable alternative to $G_{l,n}(y)$, cf. PR]. Then the explicit expressions for coverage probability error (CPE) and expected interval-width (EIW) are:

$$CPE(n, l, \beta) := P\{\theta \in I_l^\beta(\tilde{G}_{l,n}, 1)\} - \beta = CPE^*(n, l, (1+\beta)/2) - CPE^*(n, l, (1-\beta)/2),$$

where $CPE^*(n, l, \gamma) =$

$$\sum_{u=0}^{l} \sum_{v=0}^{n-l} \binom{l}{u} \binom{n-l}{v} 2^{-n} \mathbf{1}\Big[u(1+l/n) + vl/n - M(n, l, u+v, \gamma) \leq l/2\Big] - \gamma,$$

and $EIW(n, l, \beta) :=$

$$E\{width[I_l^{\beta}(\tilde{G}_{l,n}, 1)]\} = \sum_{u=0}^{n} \binom{n}{u} \Big[M(n, l, u, (1+\beta)/2) - M(n, l, u, (1-\beta)/2)\Big]/l2^{n-1},$$

where in both expressions

$$M(n, l, z, \gamma) := \min\{w : \sum_{t=max\{0, l-n+z\}}^{w} \binom{z}{t} \binom{n-z}{l-t} \geq \gamma \binom{n}{l}\}.$$

The derivations of these expressions are indicated in Appendix A.2.

Figure 1 shows the behavior of both CPE and EIW as a function of l, for fixed sample size $n = 100$ and nominal coverage $\beta = .90$. As l increases, EIW shrinks but CPE grows. Intuitively, larger l yields replicates of the point-estimator ($\tilde{s}_l^i$) that are based on more data (l) and hence tend to be closer to their target-parameter (θ), so the confidence interval which they induce is narrower. On the other hand, larger l produces fewer "independent" non-overlapping replicates (essentially n/l), so the resulting empirical quantiles ($\tilde{G}_{l,n}^{-1}(t)$) are less accurate. Another factor contributing to error (for large l) is the non-negligible *joint* distributional structure of $\tilde{s}_l^i$ with s_n: When $\tilde{s}_l^i$ comprises a substantial subset of s_n, then the random variation of $(\tilde{s}_l^i - s_n)$ (in $\tilde{G}_{l,n}(y)$) is a less appropriate empirical analog of $(s_l - \theta)$'s distribution, e.g., $Var\{\overline{X}_l^i - \overline{X}_n\}/Var\{\overline{X}_l - \theta\} = 1 - l/n$ [suggesting that $I_l^{\beta}(\tilde{G}_{l,n}, 1)$ has a tendency towards *under*coverage for large l (cf. Simulations 4.1 and 4.2, below)].

In the absence of a formal paradigm for balancing EIW versus CPE, it nevertheless looks reasonable in Figure 1 to choose l between 10 ($n^{1/2}$) and 20 ($\approx n^{2/3}$). For this range of l, there is only minor CPE (5% or less), and EIW has already dropped dramatically (by about 60%). If l increases beyond 35, then CPE becomes large enough to interfere with the interpretability of the confidence interval, and the accompanying reduction in EIW is only gradual.

4. NUMERICAL EXPERIMENTS

The Theorem (in Section 2) shows that the omnibus confidence interval asymptotically attains the correct coverage probability with interval-width shrinking to zero. This section examines the finite-sample performance of $I_l^\beta(G_{l,n}, 1)$ in two situations where constructing accurate confidence intervals is far from trivial. The first simulation involves the sample mean computed from heavy-tailed data, and the second studies a simple U-statistic [computed from serially dependent data] which can behave either as approximately Normal or as drastically non-Normal according to subtle changes in the distribution of the observations.

4.1 Example 1, Revisited: The Sample Mean Computed From Heavy-Tailed Data

The data $\{X_1, X_2, ..., X_n\}$ arise as a random sample from a heavy-tailed distribution with unknown mean θ. The statistic $s_m = \overline{X}_m$ is used to construct an omnibus confidence interval $I_l^\beta(G_{l,n}, 1)$ on θ, with nominal coverage $\beta = .90$. This procedure is repeated 1000 times for each combination of $n \in \{200, 1000, 5000\}$ and $l(n) \in \{n^{1/2}, n^{2/3}\}$, yielding the following empirical coverages and average interval-widths:

Sample Size n	Subseries Length $l(n)$	Coverage	Average Interval-Width
200	14 ($n^{1/2}$)	.890	10.00
200	34 ($n^{2/3}$)	.804	5.17
1000	32 ($n^{1/2}$)	.901	6.15
1000	100 ($n^{2/3}$)	.857	4.33
5000	71 ($n^{1/2}$)	.890	4.30
5000	292 ($n^{2/3}$)	.894	3.04

For $l(n) = n^{1/2}$, the empirical coverage is within "simulation error" of the nominal coverage at every n [note that the standard error associated with the simulated empirical coverage is approximately .01]. For $l(n) = n^{2/3}$, the empirical coverage approaches the nominal coverage as n increases. For each fixed choice of the function $l(n)$, the average interval-width shrinks as n increases. For fixed n, larger l tends to produce narrower intervals but worse coverage error $-$ echoing the tradeoffs illustrated in Section 3.1.

In this Example, computation of omnibus confidence intervals from observed data was accomplished without employing the specific underlying marginal density $p(x)$ of X_i, and without employing the asymptotic sampling distribution F or the corresponding standardizing sequence

$\{a_n\}$. This freedom from $p(x)$, F, $\{a_n\}$ is precisely the motivating strength of the omnibus confidence interval. In that light, it is now interesting to sketch the relevant theoretical analysis of the situation considered in this Example. First of all, the marginal density of X_i:

$$p(x) = \begin{cases} c_1[2log(|x|) - 1]/|x|^3, & \text{if } |x| > 3, \\ c_2, & \text{if } |x| \le 3, \end{cases}$$

where $c_1 \simeq 1.96007$ and $c_2 \simeq .086913$. This density has finite mean $\theta = 0$, but has infinite variance. Nevertheless, $p(\cdot)$ is in the "domain of attraction" of the Normal distribution, i.e., F=Normal, provided that $\{a_n\}$ is properly chosen; this follows from Ibragimov and Linnik (1971), Theorem 2.6.2, and from the fact that

$$Q(x) := P\{X_1 \ge x\} + P\{X_1 < -x\}$$

$$= (1 - 2c_2 x)\mathbf{1}\{0 < x \le 3\} + 2c_1 x^{-2} log\ x \mathbf{1}\{x > 3\} = h_Q(x)/x^2, \quad x > 0,$$

where $h_Q(\cdot)$ is slowly varying. Finally, the proper $\{a_n\}$ here is <u>not</u> $n^{1/2} \times$ *constant* (by Theorem 2.6.6 of Ibragimov and Linnik (1971)); rather, $a_n = n^{1/2}h(n)$ where $h(n)$ can be characterized via equations 2.6.12 and 2.6.15 of Ibragimov and Linnik (1971):

$$\lim_{n\to\infty} nQ\big(xn^{1/2}/h(n)\big) = 0 \text{ for all } x > 0, \text{ and}$$

$$\lim_{n\to\infty} h^2(n) \int \mathbf{1}\{|x| < \epsilon n^{1/2}/h(n)\} x^2 p(x)dx = 1 \text{ for some } \epsilon > 0.$$

<u>4.2 Example 2, Revisited: The Sample Variance Of Serially Dependent Data</u>

The data arise as n consecutive observations from a stationary time-series with unknown marginal variance θ. The statistic $s_m = \sum_{i=1}^{m}(X_i - \overline{X}_m)^2/(m-1)$ (the usual sample variance) is used to construct an omnibus confidence interval $I_l^\beta(G_{l,n}, 1)$ on θ, with nominal coverage $\beta = .90$. (The statistic s_m is a reasonable estimator because, e.g., $E\{s_m\} \to \theta$ for any stationary time-series such that $\sum_{i=1}^{\infty} Cov\{X_1, X_i\}$ converges.) This procedure is repeated 1000 times for each combination of $n \in \{200, 1000, 5000\}$ and $l(n) \in \{n^{1/2}, n^{2/3}\}$; the entire simulation experiment was carried out on time-series $\{X_i\}$, and then again on a different time-series $\{\tilde{X}_i\}$, yielding the following empirical coverages and average interval-widths:

Sample Size n	Subseries Length $l(n)$	Coverage	Average Interval-Width
$\{X_i\}$ data:			
200	14 ($n^{1/2}$)	.884	.1500
200	34 ($n^{2/3}$)	.824	.0601
1000	32 ($n^{1/2}$)	.885	.0680
1000	100 ($n^{2/3}$)	.873	.0217
5000	71 ($n^{1/2}$)	.892	.0309
5000	292 ($n^{2/3}$)	.895	.0075
$\{\tilde{X}_i\}$ data:			
200	14 ($n^{1/2}$)	.801	.251
200	34 ($n^{2/3}$)	.761	.161
1000	32 ($n^{1/2}$)	.878	.190
1000	100 ($n^{2/3}$)	.838	.105
5000	71 ($n^{1/2}$)	.890	.133
5000	292 ($n^{2/3}$)	.871	.065

Certain patterns emerge in both the $\{X_i\}$ data and the $\{\tilde{X}_i\}$ data: For each fixed choice of the function $l(n)$, empirical coverage improves and interval-width shrinks as n increases; and for fixed n, larger l tends to produce narrower intervals but worse coverage error (again echoing the tradeoffs illustrated in Section 3.1). These patterns are essentially the same as those found in Simulation 4.1 (above), even though the present simulation involves serially dependent data and involves a statistic whose asymptotic behavior is quite different (see below). Note that, for fixed (n, l), the $\{X_i\}$ intervals are much narrower than the $\{\tilde{X}_i\}$ intervals.

In this Example, computation of omnibus confidence intervals from observed data was accomplished without employing the specific underlying serial dependence mechanism, and without employing the statistic's asymptotic sampling distribution or the corresponding standardizing sequence. The relevant theoretical analysis would have been as follows: First of all, the underlying mechanism generating the data was an AR(1) process $Z_i = \rho Z_{i-1} + \xi_i$ with $\rho = .5$ and $\{\xi_i\}$ an i.i.d. sequence of standard Normal random variables; $\{X_i\}$ and $\{\tilde{X}_i\}$ were then threshold variables with $X_i = \mathbf{1}\{Z_i > 0\}$ and $\tilde{X}_i = \mathbf{1}\{Z_i > 1\}$ (such variables arise, e.g., as the outcomes in Binary Choice Models, see, e.g., Judge et al. (1985), Chapter 18). The target parameter is $\theta = .25$ for the $\{X_i\}$ data, while it is $\tilde{\theta} = .156$ for the $\{\tilde{X}_i\}$ data. Asymptotic theory shows that for the $\{X_i\}$ data, the proper standardization is $a_n = n$ and then F is the distribution with density

$$f(y) = \frac{8^{1/2}\exp((4y-1)/2\tau^2)}{\pi^{1/2}(1-4y)^{1/2}\tau}, \quad -\infty < y < 1/4, \quad \tau^2 \simeq 2.307,$$

[see Carlstein (1988), Example 4] illustrated in Figure 2; while for the $\{\tilde{X}_i\}$ data, the standardization is $\tilde{a}_n = n^{1/2}$ and $\tilde{F}$ =Normal [see Ibragimov and Linnik (1971), Theorem 18.5.4]. Thus, even though the same statistic $s_m(\cdot)$ was used on both data types ($\{X_i\}$ and $\{\tilde{X}_i\}$), there is a drastic difference between $(a_n,\ F,\ \theta)$ versus $(\tilde{a}_n,\ \tilde{F},\ \tilde{\theta})$. The omnibus confidence interval <u>automatically</u> adapts to these differences $-$ even in the presence of unknown serial dependence. In particular, the relatively narrower intervals in the $\{X_i\}$ case reflect the relatively faster rate of convergence ($a_n = n$).

APPENDIX: Proof and Derivation

A.1: Proof of the Theorem

For $t \in (0,1)$, define $A_n(t) := [s_l - G_{l,n}^{-1}(t), \infty)$. By PR's Corollary 3.2, and BPR's Lemma 1 and equation (12):

$$G_{l,n}^{-1}(t) = F^{-1}(t)/a_l + r_n(t)/a_l,$$

where $r_n(t) \xrightarrow{p} 0$ as $n \to \infty$. Therefore,

$$P\{\theta \in A_n(t)\} = P\big\{a_l(s_l - \theta) - r_n(t) \le F^{-1}(t)\big\} \overset{n\to\infty}{\to} F(F^{-1}(t)) = t.$$

The above convergence follows from R.1 $\big($recalling also that $l \to \infty$ (R.2)$\big)$ together with Slutsky's Theorem. Applying this argument for $t = (1+\beta)/2$, and similarly for $t = (1-\beta)/2$, yields the nominal asymptotic coverage probability β.

Notice that:

$$width[I_l^\beta(G_{l,n}, 1)] = \Big[F^{-1}((1+\beta)/2) - F^{-1}((1-\beta)/2) + r_n((1+\beta)/2) - r_n((1-\beta)/2)\Big]/a_l \xrightarrow{p} 0$$

as $n \to \infty$ because $a_l \to \infty$.

A.2: Derivation of CPE and EIW in Section 3.1

Let $\{j_1, j_2, ..., j_m\} \subseteq \{1, 2, ..., n\}$ and define $C(j_1, j_2, ..., j_m)$ to be the number of observed X_i's equal to "1" among $\{X_{j_1}, X_{j_2}, ..., X_{j_m}\}$; also denote $C_m := C(1, 2, ..., m)$. Then, counting shows that

$$s_l(X_{i_1}, X_{i_2}, ..., X_{i_l}) - s_n = 2(C(i_1, i_2, ..., i_l)/l - C_n/n).$$

Consider the summation in $\tilde{G}_{l,n}(y)$, for fixed y and fixed $\{X_i\}$: The summands vary only as $C(i_1, i_2, ..., i_l)$ varies, and $C(i_1, i_2, ..., i_l)$ takes the distinct values t (say) between $\max\{0, l - n + C_n\}$ and $\min\{l, C_n\}$ (inclusive) with corresponding frequencies $\binom{C_n}{t}\binom{n-C_n}{l-t}$. Thus, $\tilde{G}_{l,n}(y) = \sum_{t=max\{0,l-n+C_n\}}^{min\{l,C_n\}} \mathbf{1}\{2(t/l - C_n/n) \le y\}\binom{C_n}{t}\binom{n-C_n}{l-t}/\binom{n}{l}$ and

$$\tilde{G}_{l,n}^{-1}(\gamma) = 2\big[\min\{w : \sum_{t=max\{0,l-n+C_n\}}^{w} \binom{C_n}{t}\binom{n-C_n}{l-t} \ge \gamma\binom{n}{l}\}/l - C_n/n\big].$$

Now define

$$CPE^*(n,l,\gamma) := P\{s_l - \tilde{G}_{l,n}^{-1}(\gamma) \le 0\} - \gamma = P\{C_l - l\tilde{G}_{l,n}^{-1}(\gamma)/2 \le l/2\} - \gamma;$$

the expression for CPE^* given in Section 3.1 uses the representation $C_n = C_l + \overline{C}_l$, where $C_l \sim$ Binomial$(l, 1/2)$ and $\overline{C}_l \sim$ Binomial$(n-l, 1/2)$ are independent. The expression for EIW uses the Binomial$(n, 1/2)$ distribution of C_n and the above formula for $\tilde{G}_{l,n}^{-1}(\cdot)$.

FIGURE 1:

Coverage Probability Error and Expected Half-Width of Omnibus Confidence Interval

(Section 3.1, Example)

FIGURE 2:

Asymptotic Density $f(\cdot)$, in Section 4.2

REFERENCES

Athreya, K. (1987), "Bootstrap of the Mean in the Infinite Variance Case," in *Proceedings of the First World Congress of the Bernoulli Society* (Y. Prohorov and V. Sazonov, ed.s), 2, 95-98. VNU Science Press: The Netherlands.

Basawa, I.V., Mallik, A.K., McCormick, W.P., and Taylor, R.L. (1989), "Bootstrapping Explosive Autoregressive Processes," *Annals of Statistics*, 17, 1479-1486.

Bertail, P., Politis, D.N., and Romano, J.P. (1995), "On Subsampling Estimators With Unknown Rate of Convergence," *Technical Report #95-20*, Stanford University.

Bickel, P. and Freedman, D. (1981), "Some Asymptotic Theory for the Bootstrap," *Annals of Statistics*, 9, 1196-1217.

Bose, A. (1988), "Edgeworth Correction by Bootstrap in Autoregressions," *Annals of Statistics*, 16, 1709-1722.

Bretagnolle, J. (1983), "Lois Limites du Bootstrap de Certaines Fonctionnelles," *Annales de l'Institut Henri Poincaré*, 19, 281-296.

Carlstein, E. (1986), "The Use of Subseries Values for Estimating the Variance of a General Statistic from a Stationary Sequence," *Annals of Statistics*, 14, 1171-1179.

Carlstein, E. (1988), "Degenerate U-statistics Based on Non-independent Observations," *Calcutta Statistical Association Bulletin*, 37, 55-65.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics*, 7, 1-26.

Efron, B. (1982), "*The Jackknife, the Bootstrap and Other Resampling Plans*," SIAM: Philadelphia.

Fama, E.F. and Roll, R. (1968), "Some Properties of Symmetric Stable Distributions," *Journal of the American Statistical Association*, 63, 817-836.

Hall, P. (1982), "On Some Simple Estimates of an Exponent of Regular Variation," *Journal of the Royal Statistical Society*, B44, 37-42.

Ibragimov, I. and Linnik, Y. (1971), "*Independent and Stationary Sequences of Random Variables*," Wolters-Noordhoff: The Netherlands.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lutkepohl, H., and Lee, T.C. (1985), "*The Theory and Practice of Econometrics*," Wiley: New York.

Künsch, H. (1989), "The Jackknife and Bootstrap for General Stationary Observations," *Annals of Statistics*, 17, 1217-1241.

Politis, D.N. and Romano, J.P. (1994), "Large-Sample Confidence Regions Based on Subsamples Under Minimal Assumptions," *Annals of Statistics*, 22, 2031-2050.

Rajarshi, M.B. (1990), "Bootstrap in Markov Sequences Based on Estimates of Transition Density," *Annals of the Institute of Statistical Mathematics*, 42, 253-268.

Rosenblatt, M. (1956), "A Central Limit Theorem and a Strong Mixing Condition," *Proceedings of the National Academy of Sciences*, 42, 43-47.

Serfling, R. (1980), "*Approximation Theorems of Mathematical Statistics*," Wiley: New York.

Singh, K. (1981), "On the Asymptotic Accuracy of Efron's Bootstrap," *Annals of Statistics*, 9, 1187-1195.

Swanepoel, J. (1986), "A Note on Proving that the (Modified) Bootstrap Works," *Communications in Statistics-Theory and Methods*, 15, 3193-3203.

Wu, C.F.J. (1990), "On the Asymptotic Properties of the Jackknife Histogram," *Annals of Statistics*, 18, 1438-1452.

Young, G.A. (1994), "Bootstrap: More Than a Stab in the Dark?," *Statistical Science*, 9, 382-415.