# ASYMPTOTIC MULTIVARIATE NORMALITY FOR THE SUBSERIES VALUES OF A GENERAL STATISTIC FROM A STATIONARY SEQUENCE--WITH APPLICATIONS TO NONPARAMETRIC CONFIDENCE INTERVALS

E. Carlstein
University of North Carolina, Chapel Hill

Summary

Let $\{Z_i : -\infty < i < +\infty\}$ be a strictly stationary $\alpha$-mixing sequence with unknown marginal distributions and unknown dependence structure. Suppose that, given data $\vec{Z}_m^i = (Z_{i+1}, Z_{i+2}, \ldots, Z_{i+m})$, the statistic $s_m^i = s_m(\vec{Z}_m^i)$ is a point estimator of the unknown parameter $\theta$. If a sample series $\vec{Z}_n^0$ is available, then the subseries values $s_m^i$ $(0 \le i < i+m \le n)$ may be used to construct a nonparametric confidence interval on $\theta$ via either Student's distribution or via the Typical Value principle. The asymptotic justification for both methods rests upon a more general result regarding asymptotic multivariate normality of subseries values.

Supported by NSF Grant DMS-8400602.

Running Heading: Subseries-based Confidence Intervals

AMS 1980 Subject Classification: Primary 62G15, Secondary 62E20.

Keywords and Phrases: Confidence intervals, subseries, dependence, typical values, Student's t, jackknife, nonparametric, stationary, multivariate normality.

## 1. Introduction.

The jackknife (Tukey, 1958) and the typical-value principle (Hartigan, 1969) both employ subsample values of a general statistic as the building blocks of confidence intervals on an unknown parameter. The idea behind the jackknife is that the "pseudovalues" (which are based upon subsamples of the data) are approximately iid normal random variables; hence, they can be used to construct approximate confidence intervals based on Student's distribution. Hartigan (1975) gave sufficient conditions for the asymptotic multivariate normality of these pseudovalues, thus providing a theoretical justification for the jackknife procedure. The idea behind the typical-value principle is that certain sets of subsample values of a statistic are (approximately) "typical values" for the unknown parameter; i.e. each of the intervals between the ordered subsample values will include the unknown parameter with (approximately) equal probability. Hartigan (1975) gave sufficient conditions for a set of subsample values to behave asymptotically like a set of typical values, thus providing a theoretical justification for the concomitant confidence intervals.

All of the work discussed above deals with subsample values of a general statistic computed from iid data. If there is dependence in the data, then there is an even greater need for confidence interval techniques that free the user from parametric modeling and theoretical analysis. Indeed, the dependence structure would provide one more unknown in the model and one more complication in the theory. The present paper provides confidence interval procedures for the dependent case. The procedures are analogous to the jackknife and typical-value methods both in physical form and in spirit: "Subseries" of the data are employed rather than "subsamples" (the former referring only to subsamples composed of successive observations). On the one hand, equi-lengthed non-overlapping subseries values of a general statistic are shown to be asymptotically distributed as iid normal random variables. This justifies the use of confidence intervals based on Student's distribution, analogously to the jackknife procedure. On the other hand, certain sets of linear combinations of subseries values behave asymptotically like a set of typical values. This justifies the use of these entities in constructing confidence intervals via the typical-value principle.

The theoretical justifications that Hartigan (1975) provided for the jackknife and typical-value methods (in the case of iid data) both rested on a more general result: His Theorem 2 gave conditions under which (possibly overlapping) subsample values of a general statistic are asymptotically multivariate normal (with possibly nonzero covariances). Our theoretical justifications for confidence intervals based on Student's distribution and on the typical-value principle (under dependence) are likewise based upon a more general result: Theorem 1 (below) establishes conditions under which (possibly overlapping) subseries values of a general statistic are asymptotically multivariate normal (with possibly nonzero covariances). Quite surprisingly, the conditions in Theorem 1 are virtually identical to those in Hartigan's (1975) Theorem 2, even though the former permits dependence in the data while the latter forbids it.

The results presented below assume no knowledge of the marginal distributions of the data beyond stationarity. The only assumption about the dependence is that it be $\alpha$-mixing (a relatively weak assumption in the mixing hierarchy (see Ibragimov and Linnik, 1971)). The literature contains no other analogies to the jackknife or the typical-value methods for dependent data. Freedman (1984) has considered applying the bootstrap to a linear model with autoregressive component; but, as he emphasizes, the bootstrap calculations assume that the user has correctly specified the form of the underlying autoregressive model.

The next section presents basic notation and definitions. Section 3 contains the fundamental asymptotic multivariate normality result for subseries values of a general statistic from an $\alpha$-mixing sequence (Theorem 1). Corollaries 1 and 2 (in Section 4) provide the justifications for confidence intervals based on Student's distribution and the typical-value principle. The proof of Theorem 1 is deferred to Section 5.

## 2. Definitions and Notation.

Let $\{Z_i(\omega) : -\infty < i < +\infty\}$ be a strictly stationary sequence of real-valued random variables (r.v.) defined on probability space $(\Omega, F, P)$. Let $F_p^+$ ($F_q^-$ respectively) be the $\sigma$-field generated by $\{Z_p(\omega), Z_{p+1}(\omega), \ldots\}$ ($\{\ldots, Z_{q-1}(\omega), Z_q(\omega)\}$ respectively).

For $N \geq 1$ denote: $\alpha(N) = \sup\{|P\{A \cap B\} - P\{A\}P\{B\}| : A \in F_N^+, B \in F_0^-\}$, and define <u>$\alpha$-mixing</u> to mean $\lim_{N\to\infty} \alpha(N) = 0$.

Let $t_m(z_1, \ldots, z_m)$ be a function from $R^m \to R^1$, defined for each $m \geq 1$ so that $t_m(Z_1(\omega), \ldots, Z_m(\omega))$ is F-measurable. Suppressing the argument $\omega$ of $Z_i(\cdot)$ from here on, we denote $\vec{Z}_m^i = (Z_{i+1}, Z_{i+2}, \ldots, Z_{i+m})$ and $t_m^i = t_m(\vec{Z}_m^i)$.

E, V, and C denote expectation, variance, and covariance respectively. Indicator functions are denoted by $I\{\cdot\}$. For $B \geq 0$ denote: ${}_BX = X \cdot I\{|X| < B\}$ and ${}^BX = X - {}_BX$.

Let $\{a_n\}$ be a sequence of real vectors, and let A be a set of conditions to be satisfied by the $a_n$'s as $n \to \infty$ (e.g. $|a_n| \to \infty$). Then the notation $\lim_A x_{a_n} = x$ means that, for a single finite constant x, $\lim_{n\to\infty} x_{a_n} = x$ for <u>all</u> sequences $\{a_n\}$ satisfying A.

## 3. Main Result.

For each $n \geq 1$ the data $\vec{Z}_n^0$ from $\{Z_i\}$ is available. Consider an arbitrary but fixed number $k \geq 1$ of subseries values $\{T_{in} : 1 \leq i \leq k\}$, where $T_{in} = t_{r_{in}}^{m_{in}}$. Assume $r_{in}/n \to \rho_i^2 > 0$ as $n \to \infty$ for each i, so that none of the subseries are asymptotically negligible. Of course we must have $0 \leq m_{in} < m_{in} + r_{in} \leq n$ $\forall$ i,n so that each subseries is in fact contained in $\vec{Z}_n^0$. Also assume

$m_{in}/n \to \mu_i^2$ as $n\to\infty$ for each i, so that the proportion of overlap between subseries is asymptotically constant. In fact the asymptotic covariances between the $T_{in}$'s depend precisely upon the limiting proportions of overlap. Therefore it is convenient to pool the 2k integers $\{m_{in}, m_{in} + r_{in}: 1 \le i \le k\}$ for fixed n, and to order them and relabel them as $C_n = \{c_{in}: 1 \le i \le 2k\}$ where $0 \le c_{1n} \le c_{2n} \le \cdots \le c_{2k,n} \le n$. Thus $c_{1n} = \min_{1\le i\le k} \{m_{in}\}$ and $c_{2k,n} = \max_{1\le i\le k} \{m_{in} + r_{in}\}$. To avoid trivial complications, we assume that the ranks of $\{m_{in}, m_{in} + r_{in}: 1 \le i \le k\}$ remain the same for all n. That is, if we define $I_n(i)$ to be the rank of $m_{in}$ in $C_n$ (i.e. $c_{I_n(i),n} = m_{in}$), we may write $I_n(i) \equiv I(i)$. Similarly the rank of $m_{in} + r_{in}$ in $C_n$ may be denoted $J(i)$. Finally we define $\gamma_{i-1}^2 = \lim_{n\to\infty} (c_{in} - c_{i-1,n})/n$; in particular $\mu_i^2 = \sum_{j=I(i)}^{J(i)-1} \gamma_j^2$.

The statistics defined by $\{t_m(\cdot): m \ge 1\}$ will be called <u>central with parameter $\sigma^2$</u> if:

(I) $\lim_{A\to\infty} \overline{\lim_{n\to\infty}} A^2 P\{|t_n^0| \ge A\} = 0$; and

(II) $\lim_{A\to\infty} \overline{\lim_{n\to\infty}} A|E\{_A t_n^0\}| = 0$; and

(III) $\lim_{A\to\infty} \overline{\lim}_{u_n/w_n\to\rho^2,\, w_n\ge v_n+u_n\ge u_n\to\infty} |E\{_A t_{w_n}^0 \cdot {_A t_{u_n}^{v_n}}\} - \rho\sigma^2| = 0 \;\forall \rho^2 \in [0,1]$.

These conditions are virtually identical to those set forth by Hartigan (1975) in his definition of centrality for the independent case. Condition (I) controls the tails of $t_n^0$'s distribution; (II) centers the statistic; (III) requires the statistic to have covariance behavior analogous to the sample mean of iid r.v.s: the squared correlation between the statistic and its subseries value should be approximately equal to the proportion of overlap. Similarly to Hartigan's (1975) result for the independent case, centrality (as defined above) implies multivariate asymptotic normality of $\vec{T}_n = (T_{1n}, T_{2n}, \ldots, T_{kn})$ in the α-mixing case.

Theorem 1: Let $\{Z_i\}$ be $\alpha$-mixing. If $\{t_m(\cdot)\}$ are central with parameter $\sigma^2 > 0$, then $\vec{T}_n \xrightarrow{D} N_k(\vec{0},\Sigma)$ as $n\to\infty$. The entries of $\Sigma$ are:

$$\sigma_{ij} = \sigma^2 \sum_{\ell=\max\{I(i),I(j)\}}^{\min\{J(i),J(j)\}-1} \gamma_\ell^2/\rho_i\rho_j ,$$

where empty sums are zero.

Note that centrality does not even assume the existence of $t_m^i$'s moments. Carlstein (1985) establishes centrality for sample means, sample fractiles, and smooth functions of central statistics, all under $\alpha$-mixing. As a practical matter, the following alternative conditions (which are analogous but more restrictive) are known to imply centrality under $\alpha$-mixing (see Carlstein (1985)):

(I') $\{t_n^0\}$ are uniformly squared-integrable; and

(II') $\lim_{n\to\infty} E\{t_n^0\} = 0$; and

(III') $$\lim_{w_n \geq v_n + u_n \geq u_n \to\infty} (w_n/u_n)^{\frac{1}{2}}\, C\{t_{w_n}^0, t_{u_n}^{v_n}\} = \sigma^2 .$$

## 4. Applications to Confidence Intervals.

Throughout this section assume the following set-up: The statistic $s_m^i = s_m(\vec{Z}_m^i)$ is wholly computable from the data $\vec{Z}_m^i$, and does not depend upon any unknown parameters. Furthermore, $s_m^i$ estimates the unknown parameter $\theta$. Finally, put $t_m^i = (s_m^i - \theta)m^{\frac{1}{2}}$. The notation of Section 3 remains unchanged.

Corollary 1: Let $r_{in} = r_n$ and $m_{in} = (i-1)r_n \;\forall\, i \in \{1,\ldots,k\}$ and $\forall n$, with: $r_n/n \to \rho_o^2 > 0$, $1 \leq r_n \leq kr_n \leq n \;\forall n$. If $\{Z_i\}$ is $\alpha$-mixing and $\{t_m(\cdot)\}$ are central with parameter $\sigma^2 > 0$, then $\vec{T}_n \xrightarrow{D} N_k(\vec{0},\sigma^2 I_k)$ as $n\to\infty$.

Proof: Immediate from Theorem 1. □

Analogously to Hartigan's (1975) justification for the jackknife in the independent case, Corollary 1 provides the asymptotic justification in the $\alpha$-mixing case for treating

$$S_n = (\bar{s}_n - \theta)k^{\frac{1}{2}} / \left( \sum_{i=0}^{k-1} (s_{r_n}^{ir_n} - \bar{s}_n)^2/(k-1) \right)^{\frac{1}{2}}$$

as a r.v. with Student's distribution on k-1 degrees of freedom. (Here we denote $\bar{s}_n = \sum_{i=0}^{k-1} s_{r_n}^{ir_n}/k$.) Observe that $S_n$ is free of the nuisance parameter $\sigma^2$, and hence may be used as a pivot for constructing a confidence interval on $\theta$.

Corollary 2 shows that, even under $\alpha$-mixing, the subseries values $s_m^i$ can be used to construct a set of statistics which are (asymptotically) typical values for $\theta$.

Corollary 2: Let $m_{1n} = 0$ and $m_{i+1,n} = m_{in} + r_{in}$ $\forall i \in \{1,\ldots,k-1\}$ and $\forall n$, with: $r_{in}/n \to \rho_i^2 > 0$ $\forall i \in \{1,\ldots,k\}$, $1 \le r_{jn} < \sum_{i=1}^{k} r_{in} \le n$ $\forall j \in \{1,\ldots,k\}$ and $\forall n$. Let $\ell \in \{1,\ldots,k\}$ be arbitrary but fixed, and denote $K = \{i \in \{1,\ldots,k\}$ s.t. $i \neq \ell\}$. Define:

$$V_{in} = (s_{r_{in}}^{m_{in}}(r_{in})^{\frac{1}{2}} + s_{r_{\ell n}}^{m_{\ell n}}(r_{\ell n})^{\frac{1}{2}})/(r_{in}^{\frac{1}{2}} + r_{\ell n}^{\frac{1}{2}}) \quad \text{for each } i \in K, \forall n .$$

If $\{Z_i\}$ is $\alpha$-mixing and $\{t_m(\cdot)\}$ are central with parameter $\sigma^2 > 0$, then:

$$\lim_{n\to\infty} P\{\sum_{i\in K} I\{V_{in} < \theta\} = N\} = 1/k \quad \text{for each } N \in \{0,1,\ldots,k-1\} .$$

In particular: $P\{\min_{i\in K} V_{in} < \theta \le \max_{i\in K} V_{in}\} \to 1-2/k$ as $n\to\infty$.

Proof: By Theorem 1, $\vec{T}_n \xrightarrow{D} N_k(\vec{0},\sigma^2 I_k)$ as $n\to\infty$. For each $i \in K$, denote $\tilde{T}_{in} = T_{in} + T_{\ell n}$; also denote $\tilde{T}_n = (\tilde{T}_{in}: i \in K)$. Then $\tilde{T}_n \xrightarrow{D} N_{k-1}(\vec{0},\tilde{\Sigma})$, where

$\tilde{\Sigma} = \sigma^2(I_{k-1} + [1]_{(k-1)\times(k-1)})$. Using the logic in the second paragraph of Hartigan's (1975) proof of his Theorem 6, we may conclude that the event $A_{Nn}$ = {exactly N of the $\tilde{T}_{in}$'s, $i \in K$, are less than 0} has asymptotic probability 1/k (for each $N \in \{0,1,\ldots,k-1\}$). But since $\tilde{T}_{in} = (V_{in}-\theta)(r_{in}^{\frac{1}{2}} + r_{\ell n}^{\frac{1}{2}})$, the event $A_{Nn}$ is equivalent to the event $\sum_{i\in K} I\{V_{in}<\theta\} = N$. □

## 5. Proof of Theorem 1.

Without loss of generality, put $\sigma^2 = 1$. For each $i \in \{2,3,\ldots,2k\}$ define $\{d_{in}: n\geq 1\}$ and $\{w_{in}: n\geq 1\}$ as follows: $d_{in} = c_{in} - c_{i-1,n}$. If $\gamma_{i-1} = 0$ then $w_{in} = 0\ \forall n$; otherwise we can choose $w_{in}$ s.t. $0 \leq w_{in} \leq d_{in}\ \forall n$, $w_{in} \to \infty$, and $w_{in}/d_{in} \to 0$ as $n \to \infty$. For each $\ell \in \{1,2,\ldots,2k-1\}$ and each n, denote $\tilde{t}_{\ell n} = t^{c_{\ell n}}_{d_{\ell+1,n}-w_{\ell+1,n}}$. We shall begin by showing that:

$$B_{in} := \left|T_{in} - \sum_{\ell=I(i)}^{J(i)-1} \gamma_\ell \cdot \tilde{t}_{\ell n}/\rho_i\right| \xrightarrow{P} 0 \quad \text{as } n\to\infty \text{ for each } i \in \{1,\ldots,k\}.$$

Observe that

$$B_{in} \leq |{}^{A}T_{in}| + \sum_{\ell=I(i)}^{J(i)-1} \gamma_\ell |{}^{A}\tilde{t}_{\ell n}|/\rho_i + \left|{}_{A}T_{in} - \sum_{\ell=I(i)}^{J(i)-1} \gamma_\ell \cdot {}_{A}\tilde{t}_{\ell n}/\rho_i\right| \tag{1}$$

for $A > 0$. Let $\varepsilon > 0$ be given.

By (I), $\lim_{A\to\infty} \overline{\lim_{n\to\infty}} P\{|{}^{A}T_{in}| > \varepsilon\} = 0$ and $\lim_{A\to\infty} \overline{\lim_{n\to\infty}} P\{\gamma_\ell|{}^{A}\tilde{t}_{\ell n}| > \varepsilon\} = 0$ for each $\ell$. Let $B_{in}(A)$ denote the term within the last modulus of equation (1). We have:

$$P\{|B_{in}(A)| > \varepsilon\} \leq \varepsilon^{-2}[|E\{({}_{A}T_{in})^2\} - 1| +$$
$$\sum_{\ell=I(i)}^{J(i)-1} (2\gamma_\ell|E\{{}_{A}T_{in}\cdot {}_{A}\tilde{t}_{\ell n}\} - \gamma_\ell/\rho_i|/\rho_i + \gamma_\ell^2|E\{({}_{A}\tilde{t}_{\ell n})^2\} - 1|/\rho_i^2) +$$

$$+ 2 \sum_{I(i) \leq \ell < \ell' \leq J(i)-1} \gamma_\ell \cdot \gamma_{\ell'} |E\{_A\tilde{t}_{\ell n} \cdot {}_A\tilde{t}_{\ell' n}\}|/\rho_i^2] \,. \tag{2}$$

Now take $\lim_{A\to\infty} \overline{\lim_{n\to\infty}}$ of both sides in equation (2). Except for the terms within the last summation, each term on the r.h.s. of (2) vanishes by (III). Those remaining summands with $\gamma_\ell \cdot \gamma_{\ell'} \neq 0$ are each dominated by $|C\{_A\tilde{t}_{\ell n}, {}_A\tilde{t}_{\ell' n}\}| + |E\{_A\tilde{t}_{\ell n}\}E\{_A\tilde{t}_{\ell' n}\}|$. Now taking $\lim_{A\to\infty} \overline{\lim_{n\to\infty}}$ we obtain zero, since the covariance term is bounded by $4A^2 \alpha(c_{\ell' n} - c_{\ell+1,n} + w_{\ell+1,n})$ (see Ibragimov and Linnik, p. 306), and since (II) applies to the expectation terms. Combining these results establishes $B_{in} \overset{P}{\to} 0$ for each i.

Let $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k) \in R^k$ be fixed. It will suffice to consider the asymptotic distribution of $G_n := \sum_{i=1}^{k} \sum_{\ell=I(i)}^{J(i)-1} \lambda_i \gamma_\ell \tilde{t}_{\ell n}/\rho_i$ in place of that of $\lambda \cdot \vec{T}_n$, because $|\lambda \cdot \vec{T}_n - G_n| \leq \sum_{i=1}^{k} |\lambda_i| B_{in} \overset{P}{\to} 0$.

Write $G_n = \sum_{\ell=1}^{2k-1} \tilde{t}_{\ell n} \cdot g_\ell$, where $g_\ell = \gamma_\ell \sum_{i=1}^{k} \lambda_i \, I\{I(i) \leq \ell \leq J(i) - 1\}/\rho_i$. Define r.v.s $\{\tilde{t}'_{\ell n} : 1 \leq \ell \leq 2k-1, n \geq 1\}$ to have the same marginal distributions as $\{\tilde{t}_{\ell n} : 1 \leq \ell \leq 2k-1, n \geq 1\}$, but with $\{\tilde{t}'_{\ell n} : 1 \leq \ell \leq 2k-1\}$ independent for fixed $n \geq 1$. Let $G'_n = \sum_{\ell=1}^{2k-1} \tilde{t}'_{\ell n} \cdot g_\ell$. Then for every $u \in R$:

$$|E\{\exp\{i\,u\,G_n\}\} - E\{\exp\{i\,u\,G'_n\}\}| \leq 16 \sum_{\ell=2}^{2k-1} \alpha(w_{\ell n}) I\{\gamma_{\ell-1} \neq 0\} \to 0 \quad \text{as } n\to\infty,$$

by the argument of Ibragimov and Linnik (p. 338). Hence we may simply consider the asymptotic distribution of $G'_n$ in place of that of $\lambda \cdot \vec{T}_n$.

By Theorem 6 of Carlstein (1985), each $\tilde{t}'_{\ell n}$ is marginally asymptotically N(0,1) (provided $\gamma_\ell \neq 0$). And since $\{\tilde{t}'_{\ell n} : 1 \leq \ell \leq 2k-1\}$ are independent, we may conclude that $G'_n \overset{D}{\to} N(0, \sum_{\ell=1}^{2k-1} g_\ell^2)$. Observe that $\sum_{\ell=1}^{2k-1} g_\ell^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} \lambda_i \lambda_j \sigma_{ij}$ for $\sigma_{ij}$ as defined in Theorem 1. Since this argument holds for each $\lambda \in R^k$, the proof is completed. □

## References


Carlstein, E. (1985). Asymptotic Normality for a General Statistic from a Stationary Sequence. (To appear in Ann. Prob.)

Freedman, D. (1984). On Bootstrapping Two-Stage Least-Squares Estimates in Stationary Linear Models. Ann. Stat., 12, 827-842.

Hartigan, J. (1969). Using Subsample Values as Typical Values. JASA, 64, 1303-1317.

Hartigan, J. (1975). Necessary and Sufficient Conditions for Asymptotic Joint Normality of a Statistic and its Subsample Values. Ann. Stat., 3, 573-580.

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

Tukey, J. (1958). Bias and Confidence in Not-quite Large Samples. Ann. Math. Stat., 29, 614.