The Annals of Statistics
1984, Vol. 12, No. 1, 101–118

JACKKNIFE APPROXIMATIONS TO BOOTSTRAP ESTIMATES[1]

By Rudolf Beran

University of California, Berkeley

Let $\hat{T}_n$ be an estimate of the form $\hat{T}_n = T(\hat{F}_n)$, where $\hat{F}_n$ is the sample cdf of n iid observations and T is a locally quadratic functional defined on cdf's. Then, the normalized jackknife estimates for bias, skewness, and variance of $\hat{T}_n$ approximate closely their bootstrap counterparts. Each of these estimates is consistent. Moreover, the jackknife and bootstrap estimates of variance are asymptotically normal and asymptotically minimax. The main results: the first-order Edgeworth expansion estimate for the distribution of $n^{1/2}(\hat{T}_n - T(F))$, with F being the actual cdf of each observation and the expansion coefficients being estimated by jackknifing, is asymptotically equivalent to the corresponding bootstrap distribution estimate, up to and including terms of order $n^{-1/2}$. Both distribution estimates are asymptotically minimax. The jackknife Edgeworth expansion estimate suggests useful corrections for skewness and bias to upper and lower confidence bounds for $T(F)$.

1. Introduction. Suppose $X_1, X_2, \cdots, X_n$ are independent identically distributed random variables with unknown cdf F. Let $\hat{F}_n$ be the empirical cdf of the sample. If V is a sufficiently smooth real-valued functional defined on the set of cdf's, then $V(\hat{F}_n)$ is an asymptotically optimal estimate of $V(F)$, in the local asymptotic minimax sense. Bootstrap methods, introduced by Efron (1979), apply this familiar functional estimation idea to certain statistically interesting functionals, such as sampling distributions, which may not have closed form expressions.

Suppose that $\hat{T}_n = \hat{T}_n(X_1, X_2, \cdots, X_n)$ is an estimate of $T(F)$, where T is a specified real-valued functional. Let $H_n(x, F)$ be the cdf of $n^{1/2}[\hat{T}_n - T(F)]$. Define the standardized bias, variance, and skewness of $\hat{T}_n$ by

$$b_n(F) = n[E_F(\hat{T}_n) - T(F)], \quad s_n^2(F) = nE_F[\hat{T}_n - E_F(\hat{T}_n)]^2$$
$$k_{3,n}(F) = n^2 s_n^{-3}(F) E_F[\hat{T}_n - E_F(\hat{T}_n)]^3 \tag{1.1}$$

respectively. Then $H_n(x, \hat{F}_n)$, $b_n(\hat{F}_n)$, $s_n^2(\hat{F}_n)$ and $k_{3,n}(\hat{F}_n)$ are the respective nonparametric bootstrap estimates of the four functionals just defined. Evaluation of such bootstrap estimates is often indirect, for lack of usable closed form expressions. Possible methods of evaluation include: enumeration of all possible samples of size n from the discrete empirical distribution $\hat{F}_n$; Monte Carlo approximations based on pseudorandom samples of size n drawn from $\hat{F}_n$; and hybrid methods in which analytical simplification of the bootstrap estimate precedes evaluation by one of the first two methods. Examples and further details appear in Efron (1979).

Received May 1982; revised June 1983.
[1] Supported in part by National Science Foundation Grant MCS 80-02698.
AMS 1980 *subject classifications.* Primary 62G05; secondary 62E20.
Key words and phrases. Jackknife, bootstrap, Edgeworth expansion, asymptotic minimax.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.* www.jstor.org

Under some assumptions on $\hat{T}_n$ and T, the bootstrap distribution estimate $H_n(x, \hat{F}_n)$ is locally asymptotically minimax among all possible estimates of $H_n(x, F)$ (Beran, 1982). If F_n^* is a smoothed version of $\hat{F}_n$ such that $\| F_n^* - \hat{F}_n \| = o_p(n^{-1/2})$, where $\| \cdot \|$ denotes supremum norm, then $H_n(x, F_n^*)$ retains the local asymptotic minimax property. The further advantages and drawbacks to smoothing $\hat{F}_n$ before bootstrapping are not well understood at present.

The jackknife is an older, more specialized resampling procedure which was originally introduced by Quenouille (1956) to remove the bias of $\hat{T}_n$ and was extended by Tukey (1958) to the estimation of variance. Several subsequent authors, including Miller (1964, 1974), Brillinger (1964, 1977), and Reeds (1978) have found conditions under which the jackknife variance estimate is consistent and the bias adjusted version of $n^{1/2}[\hat{T}_n - T(F)]$ is asymptotically normal. Efron (1979) observed that the jackknife estimates of $b_n(F)$ and $s_n^2(F)$ can be viewed as analytical approximations to the bootstrap estimates $b_n(\hat{F}_n)$ and $s_n^2(\hat{F}_n)$, at least when the sample space is finite.

The principal aim of this paper is to show that the bootstrap distribution estimate $H_n(x, \hat{F}_n)$ itself may have a jackknife approximation $\hat{H}_{n,JE}(x)$ which is close enough to retain the local asymptotic minimax property of $H_n(x, \hat{F}_n)$. The basic idea is as follows. Under certain assumptions on $\hat{T}_n$, the cdf $H_n(x, F)$ has a first-order bias-corrected Edgeworth expansion:

$$H_{n,E}(x) = \Phi\left[\frac{x - n^{-1/2}b_n(F)}{s_n(F)}\right] - n^{-1/2}k_{3,n}(F)\psi\left[\frac{x - n^{-1/2}b_n(F)}{s_n(F)}\right], \tag{1.2}$$

where $\psi(x) = 6^{-1}(x^2 - 1)\varphi(x)$ and Φ, φ are, respectively, the standard normal cdf and density. Substituting jackknife estimates for the functionals $b_n(F)$, $s_n^2(F)$, $k_{3,n}(F)$ which appear on the right side of (1.2) yields a jackknife Edgeworth expansion estimate $\hat{H}_{n,JE}(x)$ for $H_n(x, F)$.

More precisely, let

$$\hat{T}_{n,i} = \hat{T}_{n+1}(X_1, X_2, \cdots, X_n, X_i), \quad \hat{T}_{n,i,j} = \hat{T}_{n+2}(X_1, X_2, \cdots, X_n, X_i, X_j) \tag{1.3}$$

and let

$$\begin{aligned} D_{n,i} &= (n+1)^2[\hat{T}_{n,i} - \hat{T}_n], \quad D_{n,i,j} = (n+2)^2[\hat{T}_{n,i,j} - \hat{T}_n] - D_{n,i} - D_{n,j} \\ \bar{D}_{n,i} &= D_{n,i} - D_{n,i,i}/2, \quad \bar{D}_{n,i,j} = D_{n,i,j} - (\bar{D}_{n,i} + \bar{D}_{n,j})/n \end{aligned} \tag{1.4}$$

for $1 \le i, j \le n$. Define the positive jackknife estimates for $b_n(F)$, $s_n^2(F)$, and $k_{3,n}(F)$ to be

$$\begin{aligned} \hat{b}_{n,J} &= n^{-1} \textstyle\sum_{i=1}^n D_{n,i}, \quad \hat{s}_{n,J}^2 = n^{-2}(n-1)^{-1} \sum_{i=1}^n \bar{D}_{n,i}^2 \\ \hat{k}_{3,n,J} &= \frac{n^{-4} \sum_{i=1}^n \bar{D}_{n,i}^3 + 3n^{-3}(n-1)^{-1} \sum\sum_{i \neq j}^{n} \bar{D}_{n,i,j}\bar{D}_{n,i}\bar{D}_{n,j}}{[n^{-3} \sum_{i=1}^n \bar{D}_{n,i}^2]^{3/2}}. \end{aligned} \tag{1.5}$$

Correspondingly, the jackknife Edgeworth expansion estimate for $H_n(x, F)$ is

$$(1.6) \qquad \hat{H}_{n,JE}(x) = \Phi\left[\frac{x - n^{-1/2}\hat{b}_{n,J}}{\hat{s}_{n,J}}\right] - n^{-1/2}\hat{k}_{3,n,J}\psi\left[\frac{x - n^{-1/2}\hat{b}_{n,J}}{\hat{s}_{n,J}}\right],$$

where ψ and Φ are as in (1.2)

Some positive jackknife estimates can exhibit severe downward bias (Hinkley, 1978). The esimates $\hat{b}_{n,J}$ and $\hat{s}^2_{n,J}$ in (1.5) are designed to handle asymptotically quadratic statistics $\hat{T}_n$. As a result, $\hat{s}^2_{n,J}$ tends to be less biased than the usual positive jackknife estimate of variance. The asymptotic results in this paper remain valid if $\hat{b}_{n,J}$ and $\hat{s}^2_{n,J}$ are replaced by more familiar negative jackknife estimates and if $\hat{k}_{3,n,J}$ is modified similarly.

It is shown in Sections 2 and 3 that $\hat{H}_{n,JE}(x)$ is asymptotically equivalent, in a certain norm which metrizes weak convergence, to the bootstrap estimate $H_n(x, d\hat{F}_n)$, up to and including terms of order $n^{-1/2}$. Consequently, $\hat{H}_{n,JE}(x)$ shares the local asymptotic mimimax property of the bootstrap estimate $H_n(x, \hat{F}_n)$. In particular, $\hat{H}_{n,JE}(x)$ dominates the normal approximation $\Phi[x/\hat{s}_{n,J}]$ and the bias-adjusted normal approximation

$$(1.7) \qquad \hat{H}_{n,JB}(x) = \Phi[(x - n^{-1/2}\hat{b}_{n,J})/\hat{s}_{n,J}].$$

These theoretical results have heuristic implications for confidence regions concerning $T(F)$. Let $c_\alpha = \Phi^{-1}(1 - \alpha)$. The form of $\hat{H}_{n,JE}(x)$ suggests

$$(1.8) \qquad \hat{T}_n - n^{-1}[\hat{b}_{n,J} + \hat{s}_{n,J}\hat{k}_{3,n,J}(c^2_\alpha - 1)/6] + n^{-1/2}\hat{s}_{n,J}c_\alpha$$

as an upper confidence bound for $T(F)$ of approximate level $1 - \alpha$. The analogous lower confidence bound is

$$(1.9) \qquad \hat{T}_n - n^{-1}[\hat{b}_{n,J} + \hat{s}_{n,J}\hat{k}_{3,n,J}(c^2_\alpha - 1)/6] - n^{-1/2}\hat{s}_{n,J}c_\alpha.$$

On the other hand, there are no apparent implications for confidence intervals based on $n^{1/2} \mid \hat{T}_n - T(F) \mid$, because the skewness and bias corrections of order $n^{-1/2}$ vanish in the implied asymptotic expansion for the cdf of $n^{-1/2} \mid \hat{T}_n - T(F) \mid$. (I am indebted to a referee for this point.)

The speculative upper and lower confidence bounds (1.8) and (1.9) receive empirical support from a Monte Carlo study which is described in Section 4. Also examined in this study are the performance of the positive jackknife estimates $\hat{b}_{n,J}$, $\hat{s}^2_{n,J}$, $\hat{k}_{3,n,J}$ and the behavior of the associated jackknife Edgeworth expansion estimate $\hat{H}_{n,JE}(x)$.

2. Asymptotics for bootstrap and jackknife estimates.

2.1 *Assumptions on $\hat{T}_n$.* Let $\mathscr{F}$ be the set of all cdf's on the real line whose support lies within a fixed compact interval I. We will suppose that the observations $\{X_i;\ 1 \le i \le n\}$ are iid and that the actual distribution of X_i has cdf belonging to $\mathscr{F}$. Let $\|\cdot\|$ denote supremum norm.

Assumption A. *The estimates* $\{\hat{T}_n; n \geq 1\}$ *are of the form* $\hat{T}_n = T(\hat{F}_n)$ *where* $\hat{F}_n$ *is the sample* cdf *and* T *is a real-valued functional defined on* $\mathscr{F}$. *The functional* T *is locally quadratic at every* F *in* $\mathscr{F}$, *in the following sense: for every* F *and* G *in* $\mathscr{F}$, *there exists a function* $t(x, y, F)$ *such that*

$$T(G) = T(F) + \int t(x, y, F)\, dG(x)\, dG(y) + r(G, F). \tag{2.1}$$

The ratio $\| G - F \|^{-2} r(G, F)$ *converges to zero as* $\| G - F \|$ *tends to zero and* $\sup\{\| G - F\|^{-2} r(G, F); G \in \mathscr{F}\} < \infty$ *for every* F *in* $\mathscr{F}$. *The function* $t(x, y, F)$ *is continuous in* (x, y) *on* I^2; $\int t(x, y, F)\, dF(x)\, dF(y) = 0$; *and* $\int [\int t(x, y, F)\, dF(y)]^2\, dF(x) > 0$. *Without loss of generality, we will assume that* $t(x, y, F)$ *is symmetrical in* x *and* y.

Some examples of estimates which satisfy Assumption A:

(a) *The* rth *sample moment* $n^{-1} \sum_{i=1}^{n} X_i^r$. Evidently, $T(F) = \int x^r\, dF(x)$ and $t(x, y, F) = 2^{-1}(x^r + y^r) - T(F)$.

(b) *The sample variance* $n^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$, where $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. In this case, $T(F) = 2^{-1} \int (x - y)^2\, dF(x)\, dF(y)$ and $t(x, y, F) = 2^{-1}(x - y)^2 - T(F)$.

(c) *L-estimates of location.* Here $T(F) = \int_0^1 F^{-1}(t) J(t)\, dt$. If J is continuously differentiable, then Assumption A holds with

$$\begin{aligned} t(x, y, F) &= -2^{-1} \int [I(x \leq z) + I(y \leq z) - 2F(z)] J \cdot F(z)\, dz \\ &\quad -2^{-1} \int [I(x \leq z) - F(z)][I(y \leq z) - F(z)] J' \cdot F(z)\, dz \end{aligned} \tag{2.2}$$

(Serfling, 1980, page 289).

(d) *M-estimates of location.* Let $\lambda_F(t) = \int \psi(x - t)\, dF(x)$, where ψ is strictly monotone with $\psi(-\infty) < 0$ and $\psi(\infty) > 0$. The functional $T(F)$ solves the equation $\lambda_F[T(F)] = 0$. If ψ is twice continuously differentiable, then Assumption A holds with

$$t(x, y; F) = \beta_F(x, y) + \beta_F(y, x) - \{2\lambda_F'[T(F)]\}^{-1} \lambda_F''[T(G)] \alpha_F(x) \alpha_F(y), \tag{2.3}$$

where

$$\begin{aligned} \alpha_F(x) &= -\{\lambda_F'[T(F)]\}^{-1} \psi[x - T(F)] \\ \beta_F(x, y) &= \alpha_F(x)[1 + \{2\lambda_F'[T(F)\}^{-1} \psi'[y - T(F)]] \end{aligned} \tag{2.4}$$

(Serfling, 1980, page 256).

2.2 *Main results.* The central concern of this paper is the asymptotic performance of jackknife and bootstrap estimates for the cdf $H_n(x, F)$ of $n^{1/2}[\hat{T}_n - T(F)]$ and for bias, variance, and skewness of $\hat{T}_n$. The principal results, obtained under Assumption A, are stated in this section; proofs are given in Section 3.

For notational convenience in what follows, let

$$t_1(x, F) = \int t(x, y, F)\, dF(y)$$
(2.6)
$$t_2(x, y, F) = t(x, y, F) - t_1(x, F) - t_1(y, F).$$

By Assumption A, $\int t_1(x, F)\, dF(x) = 0$; $t_2(x, y, F)$ is symmetrical in the arguments x, y; and $\int t_2(x, y, F)\, dF(x) = 0$. These properties imply the orthogonality relationship $\int t_1(x, F) t_2(x, y, F)\, dF(x)\, dF(y) = 0$. Evidently, $t_1(x, F) + t_1(y, F)$ is the best linear approximation to $t(x, y, F)$ in the $L_2(F \times F)$ norm.

For every F in $\mathscr{F}$, define the ball $B_n(F, c)$ as the set of distribution functions G in $\mathscr{F}$ such that $\| G - F \| \le n^{-1/2} c$. Let the notation $\sup_{n,F,c}$ designate the supremum over all distribution functions in $B_n(F, c)$. The first theorem describes asymptotic behavior of the bootstrap estimates $b_n(\hat{F}_n)$, $k_{3,n}(\hat{F}_n)$, and $s_n^2(\hat{F}_n)$.

THEOREM 1. *Suppose Assumption* A *is satisfied. Then*

$$\lim_{n\to\infty} \sup_{n,F,c} P_G[\, | b_n(\hat{F}_n) - b_n(G) | > \varepsilon] = 0$$
(2.7)
$$\lim_{n\to\infty} \sup_{n,F,c} P_G[\, | k_{3,n}(\hat{F}_n) - k_{3,n}(G) | > \varepsilon] = 0$$
$$\lim_{n\to\infty} \sup_{n,F,c} P_G[\, | s_n^2(\hat{F}_n) - s_n^2(G) | > \varepsilon] = 0$$

for every F in $\mathscr{F}$ and every positive c and ε. Moreover under every sequence of distributions $\{G_n \in B_n(F, c)\}$, the asymptotic distribution of $\{n^{1/2}[s_n^2(\hat{F}_n) - s_n^2(G_n)]\}$ is $\mathscr{N}(0, \sigma^2(F))$, where

(2.8)
$$\sigma^2(F) = 16 \int \left[t_1^2(x, F) + 2 \int t_2(x, y, F) t_1(y, F)\, dF(y) - s^2(F) \right]^2 dF(x)$$

and

(2.9)
$$s^2(F) = 4 \int t_1^2(x, F)\, dF(x).$$

The second part of Theorem 1 implies that the Lévy distance between the distribution of $n^{1/2}[s_n^2(\hat{F}_n) - s_n^2(F)]$ and its bootstrap estimate converges in probability to zero. As a result, bootstrapping the bootstrap variance estimate yields asymptotically valid confidence intervals for $s_n^2(F)$. (Use Theorem 1 in Beran (1983) and equation (3.22)).

Theorem 1 gives asymptotic normality only for $s_n^2(\hat{F}_n)$. A similar argument, under the stronger assumption that the functional T is locally cubic with remainder term of order $o(\| G - F \|^2)$, would prove asymptotic normality for $b_n(\hat{F}_n)$ and $k_{3,n}(\hat{F}_n)$.

How well does the bootstrap estimate $s_n^2(\hat{F}_n)$ estimate $s_n^2(F)$, the normalized variance of $\hat{T}_n$? With the help of Theorem 1, it is not difficult to show that $s_n^2(\hat{F}_n)$ is asymptotically minimax among all possible estimates of $s_n^2(F)$.

Let f, g denote the densities of the cdf's F, G in $\mathscr{F}$ with respect to a dominating measure μ. The Hellinger distance between F and G is defined by

$$\| G - F \|_H = \left\{ \int [g^{1/2} - f^{1/2}]^2 \, d\mu \right\}^{1/2}; \tag{2.10}$$

the choice of dominating measure μ does not affect this distance. Let $S_n(F, d)$ denote the set of distributions in $\mathscr{F}$ whose Hellinger distance from F is no greater than $n^{-1/2}d$. If $c \geq 2d$, then $S_n(F, d)$ lies within the ball $B_n(F, c)$ defined earlier.

It will be shown in Section 3.2 that

$$\lim_{n\to\infty} \sup_{G\in S_n(F,d)} n^{1/2} \left| s_n^2(G) - s_n^2(F) - 2 \int \xi(x, F) f^{1/2}(x)[g^{1/2}(x) - f^{1/2}(x)] \, d\mu \right| = 0, \tag{2.11}$$

where

$$\xi(x, F) = 4\left[t_1^2(x, F) + 2 \int t_2(x, y, F) t_1(y, F) \, dF(y) - s^2(F) \right]. \tag{2.12}$$

Note that the asymptotic variance $\sigma^2(F)$ in Theorem 1 is precisely $\int \xi^2(x, F) \, dF(x)$.

Suppose u is a non-regular monotone increasing function defined on R^+. Let $\hat{V}_n$ be any estimate of $s_n^2(F)$. Equation (2.11) and an application of the Hájek-Le Cam asymptotic minimax theorem (cf. Koshevnik and Levit, 1976) yields the following lower bound on minimax risk: for every F in $\mathscr{F}$,

$$\lim_{c\to\infty} \lim\inf_{n\to\infty} \inf_{\hat{V}_n} \sup_{n,F,c} E_G u[n^{1/2} | \hat{V}_n - s_n^2(G) |] \geq Eu[\sigma(F) | Z |], \tag{2.13}$$

where Z is a standard normal random variable and $\sigma^2(F)$ is defined by (2.8).

On the other hand, if u is also bounded, it follows from the second part of Theorem 1 that

$$\lim_{n\to\infty} \sup_{n,F,c} E_G u[n^{1/2} | s_n^2(\hat{F}_n) - s_n^2(G) |] = Eu[\sigma(F) | Z |] \tag{2.14}$$

for every positive c and every F in $\mathscr{F}$. Thus, the bootstrap variance estimate $s_n^2(\hat{F}_n)$ cannot be surpassed by any other estimate of $s_n^2(G)$ in the sense that its maximum risk over $B_n(F, c)$ is as small as possible, asymptotically in n. This result improves substantially upon an earlier version obtained by a different approach in Beran (1982).

Similarly, $s_n(\hat{F}_n)$, the bootstrap estimate of normalized standard deviation, is asymptotically minimax among all estimates of $s_n(F)$. Under every sequence $\{G_n \in B_n(F, c)\}$, the limiting distribution of $\{n^{1/2}[s_n(\hat{F}_n) - s_n(G_n)\}$ is $\mathscr{N}(0, [4s^2(F)]^{-1}\sigma^2(F))$.

The close relationship between jackknife and bootstrap estimates of bias, variance, and skewness is described in the next theorem.

THEOREM 2. *Suppose Assumption* A *is satisfied and*

$$\sup_{\|G-F\|\leq d} \sup_{x,y} | t(x, y, G) | < \infty \tag{2.15}$$

for every F in $\mathscr{F}$ and some positive d. Then

$$\lim_{n\to\infty}\sup_{n,F,c} P_G[\,|\,\hat{b}_{n,J} - b_n(\hat{F}_n)\,| > \varepsilon] = 0$$

$$\lim_{n\to\infty}\sup_{n,F,c} P_G[\,|\,\hat{k}_{3,n,J} - k_{3,n}(\hat{F}_n)\,| > \varepsilon] = 0 \tag{2.16}$$

$$\lim_{n\to\infty}\sup_{n,F,c} P_G[n^{1/2}\,|\,\hat{s}^2_{n,J} - s^2_n(\hat{F}_n)\,| > \varepsilon] = 0$$

for every F in $\mathscr{F}$ and every positive c and ε. Hence the conclusions of Theorem 1 *apply also to the jackknife estimates $\hat{b}_{n,J}$, $\hat{k}_{3,n,J}$, and $\hat{s}^2_{n,J}$.*

Because of Theorem 2, the jackknife estimate of variance is asymptotically minimax and may be bootstrapped, just as $s^2_n(\hat{F}_n)$. Investigating higher order differences between $\hat{s}^2_{n,J}$ and $s^2_n(\hat{F}_n)$ would require higher order local expansions for the functional T.

The final theorem, the main result of this paper, establishes a close relationship between the jackknifed Edgeworth expansion estimate $\hat{H}_{n,JE}(x)$ defined in (1.6) and the bootstrap estimate $H_n(x, \hat{F}_n)$. Because of the remainder term in (2.1) and because the ball $B_n(F, c)$ contains lattice distributions, convergence of the Edgeworth expansion for $\{H_n(x, G);\, G \in B_n(F, c)\}$ becomes an issue. Here we will deal with this technical problem by slightly smoothing $H_{u_n}(x, G)$.

Let v be a symmetric probability density on the real line which approximates the delta function. The relationship $\|\,d\,\|_v = \|\,d * v\,\|$, where $*$ denotes convolution, defines a semi-norm for real-valued functions d on R. If the characteristic function of v is strictly positive, then $\|\cdot\|_v$ is a norm which metrizes weak convergence. For the particular probability density

$$v(x) = \varphi(x/a), \quad a > 0 \tag{2.17}$$

whose characteristic function is $\exp[-a^2t^2/2]$, the first-order Edgeworth expansion of $H_n(x, G)$ converges locally uniformly in the norm $\|\cdot\|_v$ over all cdf's G in $B_n(F, c)$ (see Section 3 and Beran, 1982). Other choices of v are possible, the key requirement being that the characteristic function of v decay with sufficient rapidity as its argument tends to $\pm\infty$.

THEOREM 3. *Suppose Assumption* A *is satisfied,* (2.15) *holds, and v is given by* (2.17). *Then*

$$\lim_{n\to\infty}\sup_{n,F,c} P_G[n^{1/2}\,\|\,H_n(x, \hat{F}_n) - \hat{H}_{n,JE}(x)\,\|_v > \varepsilon] = 0 \tag{2.18}$$

for every F in $\mathscr{F}$ and every positive c and ε. Under every sequence $\{G_n \in B_n(F, c)\}$, the processes $\{n^{1/2}[\hat{H}_{n,JE}(x) - H_n(x, G_n)]\}$ converges weakly in $\|\cdot\|_v$ norm to the Gaussian process

$$Y_F(x) = [2s^3(F)]^{-1}\sigma(F)Zx\phi[x/s(F)], \tag{2.19}$$

where Z is a standard normal random variable.

It is immediate from this theorem that $\hat{H}_{n,JE}(x)$, like $H_n(x, \hat{F}_n)$, is an asymptotically minimax estimate of $H_n(x, F)$; see Section 2 in Beran (1982).

3. Derivations.

3.1. Approximating moments of $\hat{T}_n$. Assumption A has strong implications for the moments of $\hat{T}_n$, properties which will be used in proving the results stated in Section 2. A key fact is the following property of the empirical cdf: for every F in $\mathscr{F}$, $c > 0$, and $p \geq 1$

$$\sup_n \sup_{n,F,c} E_G[n^p \| \hat{F}_n - F \|^{2p}] \leq K_p < \infty \tag{3.1}$$

where K_p depends only on p.

To check (3.1), suppose $G_n \in B_n(F, c)$. For every $t > c$ and $n \geq 1$,

$$\begin{aligned} P_{G_n}[n^{1/2} \| \hat{F}_n - F \| \geq t] &\leq P_{G_n}[n^{1/2} \| \hat{F}_n - G_n \| \geq t - c] \\ &\leq A \exp[-(t - c)^2] \end{aligned} \tag{3.2}$$

the constant A not depending on G_n or t (Dvoretzky, Kiefer, and Wolfowitz, 1956). Since

$$E_{G_n}[n^p \| \hat{F}_n - F \|^{2p}] = \int_0^\infty P_{G_n}[n^p \| \hat{F}_n - F \|^{2p} \geq u] \, du, \tag{3.3}$$

the bound (3.1) is immediate.

LEMMA 1. *Suppose Assumption* A *is satisfied. Then, for every F in* $\mathscr{F}$ *and every* $c > 0$, $p \geq 1$

$$\lim_{n\to\infty} \sup_{n,F,c} n^p E_G \left| \hat{T}_n - T(G) - \int h(x, y, G) \, d\hat{F}_n(x) \, d\hat{F}_n(y) \right|^p = 0, \tag{3.4}$$

where

$$h(x, y, G) = t(x, y, F) - \int t(x, y, F) \, dG(x) \, dG(y). \tag{3.5}$$

Also

$$\lim_{n\to\infty} \sup_{n,F,c} n^p E_G \left| \hat{T}_n - T(G) - \int t(x, y, G) \, d\hat{F}_n(x) \, d\hat{F}_n(y) \right|^p = 0. \tag{3.6}$$

PROOF. Equation (2.1) of Assumption A implies that

$$\hat{T}_n = T(F) + \int t(x, y, F)' \, d\hat{F}_n(x) \, d\hat{F}_n(y) + r(\hat{F}_n, F). \tag{3.7}$$

Suppose $G_n \in B_n(F, c)$. By (3.2), $\{\| \hat{F}_n - F \|\}$ converges in probability to zero under $\{G_n\}$. Consequently, in view of Assumption A, the expectation under G_n of any positive power of $\| \hat{F}_n - F \|^{-2} r(\hat{F}_n, F)$ converges to zero. Combining this fact with (3.1) yields

$$\lim_{n\to\infty} n^p E_{G_n} | r(\hat{F}_n, F) |^p = 0 \tag{3.8}$$

which, with (3.7), implies

$$\lim_{n\to\infty}\sup_{n,F,c} n^p E_G \left| \hat{T}_n - T(F) - \int t(x, y, F)\, d\hat{F}_n(x)\, d\hat{F}_n(y) \right|^p = 0. \tag{3.9}$$

On the other hand, it is immediate from (2.1) and the definition of the ball $B_n(F, c)$ that

$$\lim_{n\to\infty}\sup_{n,F,c} n^p \left| T(G) - T(F) - \int t(x, y, F)\, dG(x)\, dG(y) \right|^p = 0 \tag{3.10}$$

for every $p \geq 1$. Equation (3.4) follows from (3.9) and (3.10) by Minkowski's inequality.

The derivation of (3.6) is analogous to that of (3.9), with G_n replacing F in (3.7) and (3.8). This completes the proof of the lemma.

Decompose $h(x, y, G)$ into orthogonal components

$$\begin{aligned} h_1(x, G) &= \int h(x, y, G)\, dG(y) \\ h_2(x, y, G) &= h(x, y, G) - h_1(x, G) - h_1(y, G) \end{aligned} \tag{3.11}$$

by analogy with (2.6). The next lemma describes useful asymptotic approximations to the normalized bias, variance, and skewness of $\hat{T}_n$.

LEMMA 2. *Suppose Assumption* A *is satisfied. Let*

$$b(G) = \int h(x, x, G)\, dG(x), \quad s^2(G) = 4 \int h_1^2(x, G)\, dG(x)$$

$$k_3(G) = s^{-3}(G)\left[8 \int h_1^3(x, G)\, dG(x) + 24 \int h_1(x, G) h_1(y, G) h_2(x, y, G)\, dG(x)\, dG(y)\right]. \tag{3.12}$$

Then

$$\begin{aligned} &\lim_{n\to\infty}\sup_{n,F,c} | b_n(G) - b(G) | = 0 \\ &\lim_{n\to\infty}\sup_{n,F,c} | k_{3,n}(G) - k_3(G) | = 0 \\ &\lim_{n\to\infty}\sup_{n,F,c} n^{1/2} | s_n'^2(G) - s^2(G) | = 0. \end{aligned} \tag{3.13}$$

PROOF. Setting $p = 1$ in (3.4) yields

$$\lim_{n\to\infty}\sup_{n,F,c} n \left| m_n(G) - T(G) - E_G \int h(x,y,G)\, d\hat{F}_n(x)\, d\hat{F}_n(y) \right| = 0, \tag{3.14}$$

which implies the first equation in (3.13).

Minkowski's inequality and the identity $m_n(G) = T(G) + n^{-1}b_n(G)$ give

$$\begin{aligned} &nE_G^{1/p}\left| \hat{T}_n - m_n(G) - n^{-2} \sum\sum_{i\neq j} h(X_i, X_j, G)\right|^p \\ (3.15)\qquad &\le nE_G^{1/p}\left| \hat{T}_n - T(G) - \int h(x, y, G)\, d\hat{F}_n(x)\, d\hat{F}_n(y)\right|^p \\ &\quad + E_G^{1/p} \left| n^{-1} \textstyle\sum_{i=1}^n h(X_i, X_i, G) - b_n(G)\right|^p. \end{aligned}$$

Since $h(x, y, G)$ is bounded under Assumption A, the Marcinkiewicz-Zygmund inequality implies that

$$(3.16)\qquad \sup_n \sup_{n,F,c} n^{p/2} E_G \left| n^{-1} \textstyle\sum_{i=1}^n h(X_i, X_i, G) - b_n(G)\right|^p < \infty.$$

Applying (3.4) and (3.16) to (3.15) establishes

$$(3.17)\qquad \lim_{n\to\infty} \sup_{n,F,c} nE_G^{1/p}\left| \hat{T}_n - m_n(G) - n^{-2} \sum\sum_{i\neq j} h(X_i, X_j, G)\right|^p = 0$$

for $p \ge 1$.

Let

$$(3.18)\qquad A_n(G) = n^{-2} \sum\sum_{i\neq j} h(X_i, X_j, G), \quad B_n(G) = \hat{T}_n - m_n(G)$$

and set $D_n(G) = B_n(G) - A_n(G)$. The boundedness of $t(x, y, F)$ and the orthogonal decomposition of $h(x, y, G)$ into the sum $h_1(x, G) + h_1(y, G) + h_2(x, y, G)$ yield

$$\begin{aligned} &\lim_{n\to\infty} \sup_{n,F,c} n \left| nE_G A_n^2(G) - s^2(G)\right| < \infty \\ (3.19)\qquad &\lim_{n\to\infty} \sup_{n,F,c} \left| n^2 E_G A_n^3(G) - k_3(G) s^3(G)\right| < \infty \\ &\lim_{n\to\infty} \sup_{n,F,c} n^2 E_G A_n^4(G) < \infty \end{aligned}$$

after some calculation. Since $B_n(G) = A_n(G) + D_n(G)$, we may conclude from (3.19), (3.17), and application of the Cauchy-Schwarz inequality to the cross-product terms in the expansion of $[A_n(G) + D_n(G)]^p$, that

$$\begin{aligned} &\lim_{n\to\infty} \sup_{n,F,c} n^{1/2} \left| nE_G B_n^2(G) - s^2(G)\right| = 0 \\ (3.20)\qquad &\lim_{n\to\infty} \sup_{n,F,c} \left| n^2 E_G B_n^3(G) - k_3(G) s^3(G)\right| = 0. \end{aligned}$$

The limits (3.20) imply the second and third lines in (3.13) because $s_n^2(G) = nE_G B_n^2(G)$ and $k_{3,n}(G) = n^2 E_G[B_n^3(G) s_n^{-3}(G)]$.

A variant of Lemma 2 will be needed in studying the jackknife estimates $\hat{b}_{n,J}$, $\hat{s}_{n,J}^2$, and $\hat{k}_{3,n,J}$.

LEMMA 3. *Suppose Assumption* A *is satisfied and* (2.15) *holds for every* F *in* $\mathscr{F}$ *and some positive* d. *Then the functionals* $b(G)$, $s^2(G)$, $k_3(G)$ *in* (3.13) *may be* *replaced by*

$$\bar{b}(G) = \int t(x, x, G)\, dG(x), \quad \bar{s}^2(G) = 4 \int t_1^2(x, G)\, dG(x)$$

$$\bar{k}_3(G) = \bar{s}^{-3}(G)\Big[8 \int t_1^3(x, G)\, dG(x) + 24 \int t_1(x, G)t_1(y, G)t_2(x, y, G)\, dG(x)\, dG(y)\Big] \quad (3.21)$$

The proof of this lemma strictly parallels that for Lemma 2, using (3.6) instead of (3.4). Note that the requirement (2.15) on $t(x, y, G)$ is satisfied automatically by $h(x, y, G)$ under Assumption A.

3.2. *Theorem proofs.*

PROOF OF THEOREM 1. Inequality (3.2) on the empirical cdf implies

$$\lim_{t\to\infty}\sup_n \sup_{n,F,c} P_G[\hat{F}_n \notin B_n(F, t)] = 0. \quad (3.22)$$

Combining (3.22) with Lemma 2 yields

$$\begin{aligned} &\lim_{n\to\infty}\sup_{n,F,c} P_G[\,|\, b_n(\hat{F}_n) - b(\hat{F}_n)\,| > \varepsilon] = 0 \\ &\lim_{n\to\infty}\sup_{n,F,c} P_G[\,|\, k_{3,n}(\hat{F}_n) - k_3(\hat{F}_n)\,| > \varepsilon] = 0 \\ &\lim_{n\to\infty}\sup_{n,F,c} P_G[n^{1/2}\,|\, s_n^2(\hat{F}_n) - s^2(\hat{F}_n)\,| > \varepsilon] = 0 \end{aligned} \quad (3.23)$$

for every positive ε and c. Indeed, if $C_{n,\varepsilon} = \{|\, b_n(\hat{F}_n) - b(\hat{F}_n)\,| > \varepsilon\}$, the inequality

$$P_G[C_{n,\varepsilon}] \le P_G[C_{n,\varepsilon} \cap \{\hat{F}_n \in B_n(F, t)\}] + P_G[\hat{F}_n \notin B_n(F, t)] \quad (3.24)$$

and the inclusion

$$C_{n,\varepsilon} \cap \{\hat{F}_n \in B_n(F, t)\} \subset \{\sup_{n,F,t} |\, b_n(G) - b(G)\,| > \varepsilon\} \quad (3.25)$$

imply the first line in (3.23). The other two limits in (3.23) are argued similarly.

Under $\{G_n \in B_n(F, c)\}$, the empirical product measure determined by $\hat{F}_n(x)\hat{F}_n(y)$ converges weakly, with probability one, to that determined by $F(x)F(y)$. Hence

$$\sup_{x,y} |\, h(x, y, \hat{F}_n) - h(x, y, F)\,| \to 0 \quad \text{w.p.1} \quad (3.26)$$

because $t(x, y, F)$ is a continuous, bounded function; and therefore

$$\begin{aligned} &\sup_x |\, h_1(x, \hat{F}_n) - h_1(x, F)\,| \to 0 \\ &\sup_{x,y} |\, h_2(x, y, \hat{F}_n) - h_2(x, y, F)\,| \to 0 \quad \text{w.p.1.} \end{aligned} \quad (3.27)$$

It follows from this and (3.12) that $b(\hat{F}_n)$, $s^2(\hat{F}_n)$, $k_3(\hat{F}_n)$ converge with probability one to $b(F)$, $s^2(F)$, $k_3(F)$ respectively, under every sequence $\{G_n \in B_n(F, c)\}$. Similarly, $b(G_n)$, $s^2(G_n)$, $k_3(G_n)$ also converge to $b(F)$, $s^2(F)$, $k_3(F)$ respectively.

Consequently,

$$\begin{aligned} &\lim_{n\to\infty}\sup_{n,F,c} P_G[\,|\,b(\hat{F}_n) - b(G)\,|\, > \varepsilon] = 0 \\ (3.28)\qquad &\lim_{n\to\infty}\sup_{n,F,c} P_G[\,|\,k_3(\hat{F}_n) - k_3(G)\,|\, > \varepsilon] = 0 \\ &\lim_{n\to\infty}\sup_{n,F,c} P_G[\,|\,s^2(\hat{F}_n) - s^2(G)\,|\, > \varepsilon] = 0. \end{aligned}$$

The desired locally uniform consistency result (2.7) is implied by (3.28), (3.23), and (3.13).

To prove the locally uniform asymptotic normality of $s_n^2(\hat{F}_n)$, it suffices to show, in view of (3.13) and (3.23), that the limiting distribution of $\{n^{1/2}[s^2(\hat{F}_n) - s^2(G_n)]\}$ under $\{G_n \in B_n(F, c)\}$ is $\mathcal{N}(0, \sigma^2(\mathrm{F}))$.

Clearly

$$\begin{aligned} &s^2(\hat{F}_n) \\ (3.29)\quad &= 4\int\left[\int t(x, y, F)\, d\hat{F}_n(y) - \int\int t(x, y, F)\, d\hat{F}_n(x)\, d\hat{F}_n(y)\right]^2 d\hat{F}_n(x) \\ &= A_{n,1} - A_{n,2} \end{aligned}$$

where

$$\begin{aligned} &A_{n,1} = 4\int\left[\int h(x, y, G_n)\, d\hat{F}_n(y)\right]^2 d\hat{F}_n(x) \\ (3.30)\quad & \\ &A_{n,2} = 4\left[\int\int h(x, y, G_n)\, d\hat{F}_n(x)\, d\hat{F}_n(y)\right]^2. \end{aligned}$$

A straightforward calculation shows that $E_{G_n}|\,A_{n,2}\,| = O(n^{-1})$ under Assumption A.

On the other hand,

$$\begin{aligned} (3.31)\quad A_{n,1} &= 4n^{-3} \textstyle\sum_i [\sum_j h(X_i, X_j, G_n)]^2 \\ &= 4n^{-3} \sum\sum\sum_{i\neq j\neq k} h(X_i, X_j, G_n)h(X_i, X_k, G_n) + O_p(n^{-1}), \end{aligned}$$

where the triple sum is over all triplets (i, j, k) in which no two components are equal. Replacing $h(x, y, G_n)$ by its orthogonal decomposition $h_1(x, G_n) + h_1(y, G_n) + h_2(x, y, G_n)$ yields, after some calculation,

$$\begin{aligned} (3.32)\quad A_{n,1} = &\,4\left[n^{-1} \textstyle\sum_i h_1^2(X_i, G_n) + 2n^{-2} \sum\sum_{i\neq j} h_2(X_i, X_j, G_n)h_1(X_j, G_n)\right] \\ &+ O_p(n^{-1}). \end{aligned}$$

The main term on the right side of (3.32) is a second order U-statistic, in asymmetrical form. Applying Hoeffding's well-known projection argument, we obtain

$$A_{n,1} = 4n^{-1} \sum_{i=1}^{n} \left[h_1^2(X_i, G_n) + 2 \int h_2(X_i, y, G_n) h_1(y, G_n) \, dG_n(y) \right] + O_p(n^{-1}) \tag{3.33}$$

under $\{G_n \in B_n(F, c)\}$. Thus, the limiting distribution of $\{n^{1/2}[A_{n,1} - s^2(G_n)]\}$ is $\mathcal{N}(0, \sigma^2(F))$, with $\sigma^2(F)$ defined by (2.8); note that $h(x, y, F)$ coincides with $t(x, y, F)$ under Assumption A.

In view of the previous paragraphs, the asymptotic normality of $s_n^2(\hat{F}_n)$ is now apparent. We note that the full strength of Assumption A is not needed to prove this theorem.

PROOF OF (2.11). A straightforward calculation using the definition (3.12) of $s^2(G)$ and Assumption A yields

$$s^2(G) = s^2(F) + 2 \int \xi(x, F) f^{1/2}(x)[g^{1/2}(x) - f^{1/2}(x)] \, d\mu + O(\| G - F \|_H^2), \tag{3.34}$$

for $\xi(x, F)$ as in (2.12). The limit (2.11) follows from (3.34) and (3.13).

PROOF OF THEOREM 2. Lemma 3 and (3.22) imply

$$\begin{aligned} &\lim_{n\to\infty}\sup_{n,F,c} P_G[\,| b_n(\hat{F}_n) - \bar{b}(\hat{F}_n) | > \varepsilon] = 0 \\ &\lim_{n\to\infty}\sup_{n,F,c} P_G[\,| k_{3,n}(\hat{F}_n) - \bar{k}_3(\hat{F}_n) | > \varepsilon] = 0 \\ &\lim_{n\to\infty}\sup_{n,F,c} P_G[n^{1/2} \,| s_n^2(\hat{F}_n) - \bar{s}^2(\hat{F}_n) | > \varepsilon] = 0 \end{aligned} \tag{3.35}$$

for every positive c and ε. To complete the proof of the theorem, it suffices to show that

$$\begin{aligned} &\lim_{n\to\infty}\sup_{n,F,c} P_G[\,| \hat{b}_{n,J} - \bar{b}(\hat{F}_n) | > \varepsilon] = 0 \\ &\lim_{n\to\infty}\sup_{n,F,c} P_G[\,| \hat{k}_{3,n,J} - \bar{k}_3(\hat{F}_n) | > \varepsilon] = 0 \\ &\lim_{n\to\infty}\sup_{n,F,c} P_G[n^{1/2} \,| \hat{s}_{n,J}^2 - \bar{s}^2(\hat{F}_n) | > \varepsilon] = 0 \end{aligned} \tag{3.36}$$

for every positive c and ε.

Suppose $G_n \in B_n(F, c)$ and let

$$\hat{F}_{n,i}(x) = (n+1)^{-1}[n\hat{F}_n(x) + I(X_i \le x)], \quad 1 \le i \le n. \tag{3.37}$$

From (1.4) and Assumption A,

$$D_{n,i} = (n+1)^2[T(\hat{F}_{n,i}) - T(\hat{F}_n)] = 2nt_1(X_i, \hat{F}_n) + t(X_i, X_i, \hat{F}_n) + r_{n,i} \tag{3.38}$$

where $\max\{| r_{n,i} |;\ 1 \le i \le n\} = o(1)$ because $\| \hat{F}_{n,i} - \hat{F}_n \| \le 2(n+1)^{-1}$. Thus,

$$\hat{b}_{n,J} = \int t(x, x, \hat{F}_n) \, d\hat{F}_n(x) + o_p(1) \tag{3.39}$$

under $\{G_n\}$. Let

(3.40) $\hat{F}_{n,i,j}(x) = (n+2)^{-1}[n\hat{F}_n(x) + I(X_i \le x) + I(X_j \le x)], \quad 1 \le i, \quad j \le n.$

From (1.4) and Assumption A,

$$\begin{aligned} D_{n,i,j} &= (n+2)^2[T(\hat{F}_{n,i,j}) - T(\hat{F}_n)] - D_{n,i} - D_{n,j} \\ &= 2t(X_i, X_j, \hat{F}_n) + r_{n,i,j} \end{aligned} \tag{3.41}$$

where $\max\{|r_{n,i,j}|; 1 \le i, j \le n\} = o_p(1)$ under $\{G_n\}$. Combining (3.38) with (3.41) while recalling (2.6) yields

$$\bar{D}_{n,i} = 2nt_1(X_i, \hat{F}_n) + e_{n,i} \tag{3.42}$$

and

$$\bar{D}_{n,i,j} = 2t_2(X_i, X_j, \hat{F}_n) + e_{n,i,j} \tag{3.43}$$

where $\max\{|e_{n,i}|; 1 \le i \le n\} = o_p(1)$ and $\max\{|e_{n,i,j}|; 1 \le i, j \le n\} = o_p(1)$. It follows from (1.5), (2.15), and (3.42) that

$$\hat{s}^2_{n,J} = 4 \int t_1^2(x, \hat{F}_n)\, d\hat{F}_n(x) + o_p(n^{-1}). \tag{3.44}$$

Equations (3.39) and (3.44) imply the first and third lines in (3.36). The middle line (3.36) follows from (3.42), (3.43), and the definition (1.5) of $\hat{k}_{3,n,J}$.

PROOF OF THEOREM 3. By the argument in Sections 4 and 2 of Beran (1982), Assumption A implies

$$\begin{aligned} \lim_{n\to\infty}\sup_{n,F,c} \Bigg\| & n^{1/2}\left[H_n(x, G) - \Phi\left(\frac{x}{s(G)}\right)\right] \\ & + s^{-1}(G)b(G)\phi\left(\frac{x}{s(G)}\right) + k_3(G)t\left(\frac{x}{s(G)}\right)\phi\left(\frac{x}{s(G)}\right) \Bigg\|_v = 0 \end{aligned} \tag{3.45}$$

where $t(x) = 6^{-1}(x^2 - 1)$ and Φ, ϕ are the standard normal cdf and density. Consequently,

$$\begin{aligned} \lim_{n\to\infty}\sup_{n,F,c} n^{1/2} \| & H_n(x, G) - H_n(x, F) \\ & + s^{-2}(F)[s(G) - s(F)]x\phi(x/s(F)) \|_v = 0; \end{aligned} \tag{3.46}$$

see the remarks preceding (3.28). Therefore, in view of (3.22),

$$\begin{aligned} \lim_{n\to\infty}\sup_{n,F,c} P_G[n^{1/2} \| & H_n(x, \hat{F}_n) - H_n(x, F) \\ & + s^{-2}(F)[s(\hat{F}_n) - s(F)]x\phi(x/s(F)) \|_v > \varepsilon] = 0 \end{aligned} \tag{3.47}$$

for every positive ε and c. Combining (3.46) with (3.47) yields

$$\begin{aligned} \lim_{n\to\infty}\sup_{n,F,c} P_G[n^{1/2} \| & H_n(x, \hat{F}_n) - H_n(x, G) \\ & + s^{-2}(F)[s(\hat{F}_n) - s(G)]x\phi(x/s(F)) \|_v > \varepsilon] = 0 \end{aligned} \tag{3.48}$$

for every positive ε and c. By Theorem 1, Lemma 2, and (3.23), the limiting distribution of $\{n^{1/2}[s(\hat{F}_n) - s(G_n)]\}$, where $G_n \in B_n(F, c)$, is $\mathcal{N}(0, [4s^2(F)]^{-1}\sigma^2(F))$. Hence, the processes $\{n^{1/2}[H_n(x, \hat{F}_n) - H_n(x, G_n)]\}$ converge weakly in $\|\cdot\|_v$ norm to the gaussian process $Y_F(x)$ defined in (2.19).

On the other hand, in view of Theorems 1 and 2, Lemma 2, and (1.6),

$$\text{(3.49)} \quad \begin{aligned} \lim_{n\to\infty}\sup_{n,F,c} P_G\Big[\Big\| & n^{1/2}\Big[\hat{H}_{n,JE}(x) - \Phi\Big(\frac{x}{\hat{s}_{n,J}}\Big)\Big] + \hat{s}_{n,J}^{-1}\hat{b}_{n,J}\phi\Big(\frac{x}{\hat{s}_{n,J}}\Big) \\ & + \hat{k}_{3,n,J}t\Big(\frac{x}{\hat{s}_{n,J}}\Big)\phi\Big(\frac{x}{\hat{s}_{n,J}}\Big) \Big\|_v > \varepsilon\Big] = 0 \end{aligned}$$

and therefore

$$\text{(3.50)} \quad \begin{aligned} \lim_{n\to\infty}\sup_{n,F,c} P_G\Big[\Big\| & n^{1/2}\Big[\hat{H}_{n,JE}(x) - \Phi\Big(\frac{x}{s(\hat{F}_n)}\Big)\Big] + s^{-1}(\hat{F}_n)b(\hat{F}_n)\phi\Big(\frac{x}{s(\hat{F}_n)}\Big) \\ & + k_3(\hat{F}_n)t\Big(\frac{x}{s(\hat{F}_n)}\Big)\phi\Big(\frac{x}{s(\hat{F}_n)}\Big) \Big\|_v > \varepsilon\Big] = 0 \end{aligned}$$

for every positive ε and c. The last equation draws on (2.16) and (3.23). Equations (3.45) and (3.22) entail

$$\text{(3.51)} \quad \begin{aligned} \lim_{n\to\infty}\sup_{n,F,c} P_G\Big[\Big\| & n^{1/2}\Big[H_n(x, \hat{F}_n) - \Phi\Big(\frac{x}{s(\hat{F}_n)}\Big)\Big] + s^{-1}(\hat{F}_n)b(\hat{F}_n)\phi\Big(\frac{x}{s(\hat{F}_n)}\Big) \\ & + k_3(\hat{F}_n)t\Big(\frac{x}{s(\hat{F}_n)}\Big)\phi\Big(\frac{x}{s(\hat{F})}\Big) \Big\|_v > \varepsilon\Big] = 0 \end{aligned}$$

for every positive ε and c. Combining (3.50) with (3.51) yields (2.18) and completes the proof of the theorem.

4. Numerical trial. Let T be the variance functional, so that $\hat{T}_n = T(\hat{F}_n) = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$. Suppose that F is the standard normal cdf. By classical theory,

$$\text{(4.1)} \quad b_n(F) = -1, \quad s_n^2(F) = 2(n-1)n^{-1}, \quad k_{3,n}(F) = [8n(n-1)^{-1}]^{1/2}$$

and the exact distribution of $n\hat{T}_n$ is Chi squared with $n - 1$ degrees of freedom.

The mean and standard deviation of each jackknife estimate defined in Section 1 and the levels of the associated confidence bounds for $T(F)$ were estimated by Monte Carlo methods, for sample sizes n between 10 and 80. In each case, one thousand pseudo-random $N(0, 1)$ samples of size n were used. Some of the results are summarized in the following tables.

Table 1 compares the Monte Carlo expectations of $\hat{b}_{n,J}$, $\hat{s}_{n,J}^2$, $\hat{k}_{3,n,J}$ with the population values of $b_n(F)$, $s_n^2(F)$, $k_{3,n}(F)$ respectively. The positive jackknife estimates of bias and variance are nearly unbiased for n greater than or equal to 20. However, the jackknife skewness estimate has a more persistent downward bias which diminishes slowly as n increases.

Table 2 compares the actual cdf $H_n(x, F)$ with its first-order Edgeworth expansion $H_{n,E}(x)$ [equation (1.2)] and with the Monte Carlo expectations of the jackknife estimates $\hat{H}_{n,JE}(x)$ [equation (1.6)], $\hat{H}_{n,JB}(x)$ [equation (1.7)], and $\Phi[x/\hat{s}_{n,J}]$. For n greater than or equal to 20, the Edgeworth expansion approximates $H_n(x, F)$ well over plus or minus two standard deviations. The expectation of $\hat{H}_{n,JE}(x)$ is not as close to $H_n(x, F)$, even for n equal to 40 or 80. (The downward bias in $\hat{k}_{3,n,J}$ undoubtedly contributes to this effect.) However, $\hat{H}_{n,JE}(x)$ is noticeably less biased in the center and tails than $\hat{H}_{n,JB}(x)$ and is almost uniformly less biased than the normal approximation $\Phi(x/\hat{s}_{n,J})$.

Table 3 compares the observed levels in 1,000 trials of

(a) the bias-and-skewness-corrected upper and lower confidence bounds defined in (1.8) and (1.9);

(b) the bias-corrected upper and lower confidence bounds

$$\hat{T}_n - n^{-1}\hat{b}_{n,J} \pm n^{-1/2}\hat{s}_{n,J}c_\alpha;$$

(c) the normal approximation upper and lower confidence bounds

$$\hat{T}_n \pm n^{-1/2}\hat{s}_{n,J}c_\alpha.$$

TABLE 1

Monte Carlo expectations of positive jackknife estimates for bias, variance, and skewness compared with the actual values

n	$b_n(F)$	$E(\hat{b}_{n,J})$	$s_n^2(F)$	$E(\hat{s}_{n,J}^2)$	$k_{3,n}(F)$	$E(\hat{k}_{3,n,J})$
10	−1.00	−.90	1.80	1.68	2.98	1.75
20	−1.00	−.95	1.90	1.84	2.90	2.06
40	−1.00	−.98	1.95	1.94	2.86	2.32
80	−1.00	−.99	1.98	1.98	2.85	2.52

TABLE 2

Monte Carlo expectations of three jackknife cdf *estimates compared with the actual* cdf $H_n(x, F)$ *and with the actual Edgeworth expansion* $H_{n,E}(x)$

	x	$H_n(x, F)$	$H_{n,E}(x)$	$E[\hat{H}_{n,JE}(x)]$	$E[\hat{H}_{n,JB}(x)]$	$E[\Phi(x/\hat{s}_{n,J})]$
	−3.0	.0083	.0090	.017	.025	.020
	−1.5	.1675	.1697	.149	.149	.125
	−.75	.3587	.3584	.331	.314	.274
$n = 40$	0	.5744	.5746	.571	.547	.500
	.75	.7567	.7563	.773	.763	.726
	1.5	.8788	.8764	.892	.896	.875
	3.0	.9790	.9785	.978	.985	.981
	−3.0	.0113	.0117	.016	.022	.019
	−1.5	.1607	.1619	.152	.152	.135
	−.75	.3400	.3400	.328	.315	.287
$n = 80$	0	.5526	.5527	.551	.532	.500
	.75	.7413	.7411	.749	.740	.713
	1.5	.8721	.8708	.879	.882	.865
	3.0	.9798	.9795	.979	.984	.981

TABLE 3
Estimated levels of three jackknife upper and lower confidence bounds for $T(F)$

	Nominal Level	Estimated Level					
		Bias-and-skewness adjusted bounds		Bias-adjusted bound		Normal approximation bound	
		Upper	Lower	Upper	Lower	Upper	Lower
	.975	.988	.983	.995	.975	.995	.983
	.95	.973	.964	.980	.954	.976	.968
	.90	.921	.921	.932	.910	.914	.929
$n = 40$	.85	.876	.872	.877	.872	.857	.894
	.80	.824	.815	.815	.817	.793	.853
	.70	.711	.714	.693	.732	.660	.764
	.60	.613	.622	.593	.636	.540	.675
	.975	.974	.976	.986	.969	.981	.975
	.95	.951	.948	.958	.942	.955	.951
	.90	.903	.901	.906	.899	.897	.908
$n = 80$	.85	.852	.859	.853	.859	.835	.871
	.80	.805	.809	.801	.814	.782	.835
	.70	.701	.718	.687	.727	.670	.751
	.60	.607	.614	.591	.632	.567	.666

For n equal to 40 and 80, the bias-and-skewness-corrected confidence bounds are the most reliable, particularly when the nominal level is close to 1 or to .5. For smaller n, the discrepancies between nominal levels and observed levels are unsatisfactory, no matter which confidence bounds are used.

REFERENCES

BERAN, R. J. (1982). Estimated sampling distributions: the bootstrap and competitors. *Ann. Statist.* **10** 212–225.

BERAN, R. J. (1983). Bootstrap methods in statistics. *Jber. d. Dt. Math.-Verein.* **85**, to appear.

BICKEL, P. J. and FREEDMAN, D. (1980). On Edgeworth expansions for the bootstrap. Unpublished manuscript.

BRILLINGER, D. R. (1964). The asymptotic behavior of Tukey's general method of setting approximate confidence limits (the jackknife) when applied to maximum likelihood estimates. *Rev. Inst. Internat. Statist.* **32** 202–206.

BRILLINGER, D. R. (1977). Approximate estimation of the standard errors of complex statistics based on sample surveys. *New Zealand Statistician* **11** 35–41.

DVORETZKY, A. KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

EFRON, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7** 1–26.

HINKLEY, D. V. (1978). Improving the jackknife with special reference to correlation estimation. *Biometrika* **65** 13–21.

KOSHEVNIK, YU. A. and LEVIT, B. YA. (1976). On a nonparametric analog of the information matrix. *Theory Probab. Appl.* **21** 738–753.

MILLER, R. G. (1964). A trustworthy jackknife. *Ann. Math. Statist.* **35** 1594–1605.

MILLER, R. G. (1974). The jackknife: a review. *Biometrika* **61** 1–15.

QUENOUILLE, M. H. (1956). Notes on bias in estimation. *Biometrika* **43** 353–360.

REEDS, J. A. (1978). Jackknifing maximum likelihood estimates. *Ann. Statist.* **6** 727–739.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.
SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.
TUKEY, J. W. (1958). Bias and confidence in not-quite large samples. *Ann. Math. Statist.* **29** 614.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720