The Annals of Statistics
1982, Vol. 10, No. 1, 212–225

ESTIMATED SAMPLING DISTRIBUTIONS: THE BOOTSTRAP AND COMPETITORS[1]

By Rudolf Beran

University of California, Berkeley

Let $X_1, X_2, \cdots, X_n$ be i.i.d random variables with d.f. F. Suppose the $\{\hat{T}_n = \hat{T}_n(X_1, X_2, \cdots, X_n); n \geq 1\}$ are real-valued statistics and the $\{T_n(F); n \geq 1\}$ are centering functionals such that the asymptotic distribution of $n^{1/2}\{\hat{T}_n - T_n(F)\}$ is normal with mean zero. Let $H_n(x, F)$ be the exact d.f. of $n^{1/2}\{\hat{T}_n - T_n(F)\}$. The problem is to estimate $H_n(x, F)$ or functionals of $H_n(x, F)$. Under regularity assumptions, it is shown that the bootstrap estimate $H_n(x, \hat{F}_n)$, where $\hat{F}_n$ is the sample d.f., is asymptotically minimax; the loss function is any bounded monotone increasing function of a certain norm on the scaled difference $n^{1/2}\{H_n(x, \hat{F}_n) - H_n(x, F)\}$. The estimated first-order Edgeworth expansion of $H_n(x, F)$ is also asymptotically minimax and is equivalent to $H_n(x, \hat{F}_n)$ up to terms of order $n^{-1/2}$. On the other hand, the straightforward normal approximation with estimated variance is usually not asymptotically minimax, because of bias. The results for estimating functionals of $H_n(x, F)$ are similar, with one notable difference: the analysis for functionals with skew-symmetric influence curve, such as the mean of $H_n(x, F)$, involves second-order Edgeworth expansions and rate of convergence n^{-1}.

1. Introduction. Let $X_1, X_2, \cdots, X_n$ be independent identically distributed random variables with unknown distribution function F. Suppose the $\{\hat{T}_n = \hat{T}_n(X_1, X_2, \cdots, X_n); n \geq 1\}$ are real-valued statistics and the $\{T_n(F); n \geq 1\}$ are real-valued functionals such that the asymptotic distribution of $n^{1/2}\{\hat{T}_n - T_n(F)\}$ is normal with mean zero. Let $H_n(x, F)$ be the exact distribution function of $n^{1/2}\{\hat{T}_n - T_n(F)\}$. A basic problem in statistics is the estimation of $H_n(x, F)$ or functionals of $H_n(x, F)$ from the sample. Indeed, the mean and variance of $H_n(x, F)$ are, respectively, the bias and variance of $\hat{T}_n$ when $\hat{T}_n$ is regarded as an estimate of $T_n(F)$. Moreover, confidence intervals for $T_n(F)$ can be constructed from a knowledge of $H_n(x, F)$.

Possible estimates of $H_n(x, F)$ include Edgeworth expansions of various orders with coefficients estimated from the data; and the bootstrap estimate $H_n(x, \hat{F}_n)$, where $\hat{F}_n$ is the sample d.f. Since exact calculation of $H_n(x, \hat{F}_n)$ may be difficult, Monte Carlo approximations are often used (Efron, 1979). A functional of $H_n(x, F)$ may be estimated directly from an estimate of $H_n(x, F)$; or by other methods, such as the jackknife, suggested by expansions for $n^{1/2}\{\hat{T}_n - T_n(F)\}$; c.f. Brillinger (1977) for a discussion of variance estimates.

If $H_n(x, F)$ depends smoothly upon F, it is plausible that the bootstrap estimate $H_n(x, \hat{F}_n)$ will be consistent in the following sense: $H_n(x, \hat{F}_n)$ converges weakly, in probability, to the same normal distribution as does $H_n(x, F)$. Consistency of the bootstrap estimate has been proved under various assumptions by Efron (1979) and by Bickel and Freedman (1981). The latter paper shows that the convergence is typically almost sure.

Less apparent to the intuition are the asymptotic distributions of bootstrap estimates and answers to the following questions:

(a) How well can $H_n(x, F)$ or a functional of $H_n(x, F)$ be estimated?

Received January 1981; revised September 1981.

[1] This research was supported in part by National Science Foundation Grants MCS 75-10376 and MCS 80-02698.

AMS 1980 *subject classification*. Primary 62G05; secondary 62E20.

Key words and phrases. Sampling distribution, bootstrap estimates, jackknife, asymptotic minimax, Edgeworth.

(b) What are (asymptotically) optimal estimates for $H_n(x, F)$ and for functionals of $H_n(x, F)$?

These matters are the subject of this paper.

Section 2 of the paper contains results which characterize asymptotically minimax estimates of $H_n(x, F)$. Under certain assumptions on the statistics $\{\hat{T}_n;\ n \geq 1\}$, the bootstrap estimate $H_n(x, \hat{F}_n)$ and the first-order Edgeworth expansion estimate are both asymptotically minimax. The empirical processes $n^{1/2}\{H_n(x, \hat{F}_n) - H_n(x, F)\}$ converge weakly to a degenerate Gaussian process. The same limit process occurs for the first-order Edgeworth expansion estimate. The results for estimating functionals of $H_n(x, F)$ are similar, with one notable difference: the analysis for functionals with skew-symmetric influence curve, such as the mean of $H_n(x, F)$, involves second-order Edgeworth expansions and rate of convergence n^{-1}.

The principal assumption underlying the results described above is, roughly, that the Edgeworth expansion for $H_n(x, G)$ hold uniformly over all distributions G in certain shrinking balls about F; the radius of each ball is of order $n^{-1/2}$. Potential technical difficulties with lattice distributions are avoided by slightly smoothing $H_n(x, G)$ in the loss functions considered. The existence of locally uniform Edgeworth expansions for smoothed $H_n(x, G)$ is discussed in Section 3.

It should not be imagined, however, that bootstrap estimates behave reasonably for every sequence of statistics $\{\hat{T}_n\}$ whose asymptotic distribution is normal. Suppose that $F(x) = \Phi(x - \theta)$, Φ being the standard normal distribution function, and that $\hat{T}_n$ is the following estimate (proposed by J. L. Hodges in a different context; cf. LeCam, 1953):

$$\hat{T}_n = \begin{cases} b\bar{X}_n & \text{if } |\bar{X}_n| \leq n^{-1/4}, \\ \bar{X}_n & \text{otherwise,} \end{cases} \tag{1.1}$$

where $\bar{X}_n$ is the sample average and b is any positive constant. Let $T_n(F)$ be the mean of F. Then $\{H_n(x, F)\}$ converges uniformly to $\Phi(b^{-1}x)$ if $\theta = 0$ and to $\Phi(x)$ if $\theta \neq 0$. On the other hand, the bootstrap estimates $\{H_n(x, \hat{F}_n)\}$ converge uniformly to $\Phi(x)$ w.p.1 if $\theta \neq 0$; but

$$\sup_x |H_n(x, \hat{F}_n) - \Phi(b^{-1}\{x - n^{1/2}\bar{X}_n(b-1)\})| \to 0 \qquad \text{w.p.1} \tag{1.2}$$

if $\theta = 0$. The bootstrap estimates are not consistent if b differs from 1.

The asymptotic behavior of $H_n(x, \hat{F}_n)$ in this example can be derived as follows. Let $X_1^*, X_2^*, \cdots, X_n^*$ be i.i.d. random variables whose conditional distribution, given $X_1, X_2, \cdots, X_n$, is $\hat{F}_n$. Let $\hat{T}_n^*$ denote the Hodges estimate calculated from the $\{X_i^*\}$. Then

$$H_n(x, \hat{F}_n) = P[n^{1/2}\{\hat{T}_n^* - T_n(\hat{F}_n)\} \leq x \mid X_1, X_2, \cdots, X_n]. \tag{1.3}$$

If $\theta = 0$,

$$\begin{aligned} &P(|\bar{X}_n^*| \leq n^{-1/4} \mid X_1, X_2, \cdots, X_n) \\ &\quad = P\{-n^{1/4} - n^{1/2}\bar{X}_n \leq n^{1/2}(\bar{X}_n^* - \bar{X}_n) \leq n^{1/4} - n^{1/2}\bar{X}_n \mid X_1, X_2, \cdots, X_n\}. \end{aligned} \tag{1.4}$$

With probability one, the conditional distribution of $n^{1/2}(\bar{X}_n^* - \bar{X}_n)$, given $X_1, X_2, \cdots, X_n$, converges weakly to the standard normal distribution. Also $\pm n^{1/4} - n^{1/2}\bar{X}_n$ converges with probability one to $\pm\infty$ because $n^{1/2}\bar{X}_n$ has a $N(0, 1)$ distribution. (For instance, if $A_n = \{n^{1/4} - n^{1/2}\bar{X}_n < n^{1/8}\}$, then $P(A_n) < Cn^{-2}$ for all sufficiently large n, C being a constant. Hence, by the Borel-Cantelli lemma, $P(A_n$ occurs infinitely often$) = 0$.) These last two facts imply that

$$\lim_{n\to\infty} P(|\bar{X}_n^*| \leq n^{1/4} \mid X_1, X_2, \cdots, X_n) = 1 \qquad \text{w.p.1.} \tag{1.5}$$

Equivalently, using the definition of the Hodges estimate and the fact that T_n is the mean functional,

$$\lim_{n\to\infty} P[n^{1/2}\{\hat{T}_n^* - T_n(\hat{F}_n)\} \tag{1.6}$$
$$= bn^{1/2}(\bar{X}_n^* - \bar{X}_n) + n^{1/2}\bar{X}_n(b-1) \mid X_1, \cdots, X_n] = 1 \quad \text{w.p.1.}$$

Equation (1.6) implies (1.2) because the conditional distribution function of $n^{1/2}(\bar{X}_n^* - \bar{X}_n)$ converges uniformly to $\Phi(x)$ w.p.1. If $\theta \neq 0$, the analysis is similar, the key fact in this case being

$$\lim_{n\to\infty} P(\hat{T}_n^* = \bar{X}_n^* \mid X_1, X_2, \cdots, X_n) = 1 \quad \text{w.p.1.} \tag{1.7}$$

The main point of the example is this: the bootstrap is not foolproof, even for statistics $\{\hat{T}_n\}$ whose asymptotic distribution is normal. Asymptotic optimality, or even consistency, of the bootstrap estimate $H_n(x, \hat{F}_n)$ is not to be expected unless $H_n(x, F)$ depends smoothly upon F. For examples of bootstrap failure in cases where the limit law is not normal, see Bickel and Freedman (1981).

The estimation theory for sampling distributions other than $H_n(x, F)$ is also of interest. For instance:

(a) Let $L_n(x, F)$ be the distribution function of $n^{1/2}\{\hat{T}_n - T_n(F)\}/s_n(F)$, where $s_n(F)$ is a scaling functional chosen so that $L_n(x, F)$ converges weakly to the standard normal distribution. The asymptotic estimation theory for $L_n(x, F)$ parallels that for $H_n(x, F)$, though the rate of convergence becomes n^{-1} rather than $n^{-1/2}$ if the centering functionals $T_n(F)$, $s_n(F)$ are chosen carefully (i.e. as in Assumption 1′ of Section 2.3). The approach used in Section 2.1 still works, with Assumption 1′ in place of Assumption 1 and other fairly obvious modifications; the bootstrap estimate $L_n(x, \hat{F}_n)$ is asymptotically minimax for $L_n(x, F)$. Despite the better rate of convergence, statistical applications for $L_n(x, \hat{F}_n)$ seem few. Typically, the value of $s_n(F)$ is not known at the unknown true F.

(b) Let $K_n(x, F)$ be the distribution function of $n^{1/2}\{\hat{T}_n - T_n(F)\}/\hat{s}_n$, where $\hat{s}_n$ is a consistent estimate of $s_n(F)$. Like $H_n(x, \hat{F}_n)$, $K_n(x, \hat{F}_n)$ may be used to construct confidence intervals for $T_n(F)$. For the simplest example, namely the t-statistic, the method of Section 3.2 yields a first order asymptotic expansion for $K_n(x, F)$ which satisfies Assumption 1. This implies that the rate of convergence of $K_n(x, \hat{F}_n)$ to $K_n(x, F)$ is better than $n^{-1/2}$, in the norm $\|\cdot\|_v$ defined in Section 2.1. Details of the argument will be given elsewhere. We conjecture that the rate of convergence is n^{-1} and that $K_n(x, \hat{F}_n)$ is asymptotically minimax for $K_n(x, F)$. Checking this conjecture would require a second order expansion for $K_n(x, F)$.

2. Asymptotically optimal estimates.

2.1 *Estimating the sampling distribution* $H_n(x, F)$. Let μ be a sigma-finite measure on the real line. If h is a bounded function and $k \in L^1(\mu)$, let $\langle k, h\rangle = \int kh\, d\mu$ and let $\|h\| = \sup_x |h(x)|$. Let $\mathscr{F}$ denote the set of distribution functions F considered possible for the data. For every F in $\mathscr{F}$, define the ball $B_n(F, c)$ as the set of distribution functions G such that $\|G - F\| \le n^{-1/2}c$. Roughly speaking, we will assume that $H_n(x, G)$, the d.f. of $n^{1/2}\{\hat{T}_n - T_n(G)\}$ under G, has a first-order Edgeworth expansion which holds uniformly over all distributions G in $B_n(F, c)$; the coefficients of the expansion are to be smooth functionals of G; both assertions are to hold for every positive c and every d.f. F in $\mathscr{F}$. The latter requirement ensures that the optimality results of this section hold for the unknown true F responsible for the data.

Since the ball $B_n(F, c)$ contains lattice distributions, the Edgeworth expansion assumption needs to be formulated more carefully. Let $v(x)$ be a symmetric probability density on the real line which approximates the delta function; for instance

$$v(x) = a^{-1}(1 - a^{-1}|x|) \quad \text{if} \quad |x| \le a. \tag{2.1}$$

Let $H_{n,v}(x, G)$ be the convolution of $H_n(x, G)$ with $v(x)$,

$$H_{n,v}(x, G) = \int H_n(x - y, G)v(y)\, dy. \tag{2.2}$$

Similarly, let Φ_v and φ_v denote the convolution of v with the standard normal d.f. and density, respectively. The subscript v will be used routinely to indicate the convolution of the subscripted function with v.

Let the notation $\sup_{n,F,c}$ designate the supremum over all distribution functions G in $B_n(F, c)$. The precise statement of the Edgeworth expansion assumption is as follows.

ASSUMPTION 1. Let $J_n(x, G)$ be the distribution function of $n^{1/2}\{T_n - m_n(G)\}/s_n(G)$ under G, where $m_n(G)$ and $s_n(G)$ are appropriately chosen centering functionals. The following assertions hold for every positive c and every F in $\mathscr{F}$:

(a) $J_{n,v}(x, G)$ has a first order Edgeworth expansion

$$\lim_{n\to\infty}\sup_{n,F,c}\| n^{1/2}\{J_{n,v}(x, G) - \Phi_v(x)\} + k(G)t_v(x)\| = 0, \tag{2.3}$$

where $k(G)$ is a functional of G and $t(x) = 6^{-1}(x^2 - 1)\varphi(x)$. The subscript v indicates convolution with v.

(b) There exists a functional $b(G)$ such that

$$\lim_{n\to\infty}\sup_{n,F,c} n\,| m_n(G) - T_n(G) - n^{-1}b(G)| = 0. \tag{2.4}$$

(c) The functionals $\{s_n(G)\}$ are differentiable at F: there exists s'_F in $L^1(\mu)$ such that

$$\lim_{n\to\infty}\sup_{n,F,c}| n^{1/2}\{s_n(G) - s_n(F)\} - n^{1/2}\langle s'_F, G - F\rangle| = 0. \tag{2.5}$$

Moreover, $s(F) = \lim_{n\to\infty} s_n(F)$ exists.

(d) The functionals $b(G)$ and $k(G)$ are $\|\cdot\|$ continuous at F.

To clarify this assumption, consider the following example. Let $\hat{T}_n$ be the sample mean and let $\mathscr{F}$ be the set of all distribution functions supported on a fixed compact interval. A weaker form of Assumption 1, in which $B_n(F, c)$ is replaced by $B_n(F, c) \cap \mathscr{F}$, holds in this case. Indeed, assertion (a) is true provided $m_n(G)$, $s_n(G)$, $k(G)$ are respectively the mean of G, the standard deviation of G, and the standardized third cumulant of G. (This follows by examination of the classical Edgeworth expansion argument for the sample mean in the nonlattice case.) Assertion (b) is trivial for the natural choice $T_n(G) =$ mean of $G = m_n(G)$. Assertions (c) and (d) hold because G lies in $\mathscr{F}$ (integrate by parts). This weaker variant of Assumption 1 suffices for the results of this section because $\mathscr{F}$ is large enough to contain $\hat{F}_n$ with probability one. Section 3 gives examples of statistics $\{\hat{T}_n\}$ which satisfy Assumption 1 unmodified.

The relationship between $H_{n,v}(x, G)$ and $J_{n,v}(x, G)$ is

$$H_{n,v}(x, G) = J_{n,v}(s_n^{-1}(G)[x - n^{1/2}\{m_n(G) - T_n(G)\}], G). \tag{2.6}$$

It follows from Assumption 1 that for every positive c

$$\begin{aligned}\lim_{n\to\infty}\sup_{n,F,c}\| n^{1/2}\{&H_{n,v}(x, G) - \Phi_v(x/s_n(G))\}\\ &+ k(G)t_v(x/s_n(G)) + s_n^{-1}(G)b(G)\varphi_v(x/s_n(G))\| = 0.\end{aligned} \tag{2.7}$$

Consequently,

$$\lim_{n\to\infty}\sup_{n,F,c}\| n^{1/2}\{H_{n,v}(x, G) - H_{n,v}(x, F)\} - n^{1/2}\langle h_F, G - F\rangle w_F(x)\| = 0, \tag{2.8}$$

where

$$h_F(x) = -s'_F(x)/s^2(F), \qquad w_F(x) = x\varphi_v(x/s(F)). \tag{2.9}$$

The relationship $\|f\|_v = \|f_v\|$ for all functions f on the real line defines a semi-norm $\|\cdot\|_v$. If the characteristic function of v is strictly positive almost everywhere, then $\|\cdot\|_v$ is a norm.

Let $\hat{H}_n(x)$ be any estimate of $H_n(x, G)$. The risk of $\hat{H}_n$ will be defined as

$$\begin{aligned}R_n(\hat{H}_n, G) &= E_G u(n^{1/2}\|\hat{H}_n(x) - H_n(x, G)\|_v)\\ &= E_G u(n^{1/2}\|\hat{H}_{n,v}(x) - H_{n,v}(x, G)\|),\end{aligned} \tag{2.10}$$

where $\hat{H}_{n,v}$ is the convolution of $\hat{H}_n$ with v and $u: R^+ \to R^+$ is monotone increasing. The smoothing by convolution with v in measuring the distance between $\hat{H}_n(x)$ and $H_n(x, G)$ is a technical device; without smoothing Assumption 1 fails to hold because $B_n(F, c)$ contains lattice distributions. Whether optimality results resembling those obtained in this paper can be established without smoothing is an open question.

Let

$$h_{F,0}(x) = \int_x^\infty h_F(y)\, d\mu(y) - \int\int_x^\infty h_F(y)\, d\mu(y)\, dF(x). \tag{2.11}$$

THEOREM 1. *Suppose Assumption 1 is satisfied. Then, for every F in $\mathscr{F}$,*

$$\lim_{c\to\infty} \liminf_{n\to\infty} \inf_{\hat{H}_n} \sup_{n,F,c} R_n(\hat{H}_n, G) \geq R_0(F), \tag{2.12}$$

where

$$R_0(F) = \int u(a\,|\,z\,|)\varphi(z)\, dz \tag{2.13}$$

and

$$a = \|w_F\| \cdot \left\{\int h_{F,0}^2(x)\, dF(x)\right\}^{1/2}. \tag{2.14}$$

PROOF. Let $e_F(x) = \{\int h_{F,0}^2\, dF\}^{-1/2} h_{F,0}(x)$. For every real t, let $\{G_{n,t};\ n \geq 1\}$ be a sequence of d.f.'s such that

$$dG_{n,t}/dF = 1 + n^{-1/2} t e_F \tag{2.15}$$

for all sufficiently large n (e_F is bounded). Since w_F is continuous, vanishing at $\pm\infty$, there exist a probability λ and functions $\{d_i;\ i \geq 1\}$ of unit length in $L^1(\lambda)$ such that $\|w_F\| = \sup\{|\langle d_i, w_F\rangle_0|;\ i \geq 1\}$; here $\langle \cdot, \cdot \rangle_0$ is the inner product with respect to λ.

This last claim can be justified as follows. Let λ be any probability on R whose d.f. M is continuous and strictly monotone. Observe that $\|w_F\| = \sup\{|\langle d, w_F\rangle_0| : d$ of unit length in $L^1(\lambda)\}$ and that $\langle d, w_F\rangle_0$ equals the inner product in $L^1[0, 1]$ *of* $d \cdot M^{-1}$ and $w_F \cdot M^{-1}$. Let $\{d_i^* : i \geq 1\}$ be a countable dense subset of the unit ball in $L^1[0, 1]$. Define $d_i(x) = d_i^* \cdot M(x)$ for every value of i.

Without loss of generality, assume that the loss function u is bounded and continuous as well as monotone, hence is uniformly continuous on R^+. Note that $\lim_n n^{1/2} \|G_{n,t} - F\| \leq 2^{-1} A^{-1} |t|$, where $A^{-1} = 2\|e_F\|$. For every positive c,

$$\begin{aligned} &\liminf_{n\to\infty} \inf_{\hat{H}_n} \sup_{n,F,c} R_n(\hat{H}_n, G) \\ &\geq \liminf_{n\to\infty} \inf_{\hat{H}_n} \sup\{E_{G_{n,t}} u(\|n^{1/2}\{\hat{H}_{n,v} - H_{n,v}(\cdot, F)\} - t a_0 w_F\|);\ |t| \leq Ac\} \\ &= \liminf_{n\to\infty} \inf_{Y_n} \sup\{E_{G_{n,t}} u(\|Y_n - t a_0 w_F\|);\ |t| \leq Ac\}, \end{aligned} \tag{2.16}$$

where $a_0 = \{\int h_{F,0}^2\, dF\}^{1/2}$ and $Y_n = n^{1/2}\{\hat{H}_{n,v} - H_{n,v}(\cdot, F)\}$. The second-to-last step depends on (2.15), (2.8) and the definition of e_F.

Moreover, for every $k \geq 1$,

$$\begin{aligned} &\lim_{c\to\infty} \liminf_{n\to\infty} \inf_{Y_n} \sup_{|t|\leq c} E_{G_{n,t}} u(\|Y_n - t a_0 w_F\|) \\ &\geq \lim_{c\to\infty} \liminf_{n\to\infty} \inf_{Y_n} \sup_{|t|\leq c} E_{G_{n,t}} u(\max_{i\leq k} |\langle d_i, Y_n\rangle_0 - t a_0 \langle d_i, w_F\rangle_0|) \\ &\geq \int u(a_0\,|\,z\,|\max_{i\leq k} |\langle d_i, w_F\rangle_0|)\varphi(z)\, dz. \end{aligned} \tag{2.17}$$

The final inequality in (2.17) is a well-known variant of Hájek's (1972) asymptotic minimax theorem; its justification rests upon the fact that

$$P_F(|\textstyle\sum_{i=1}^n \log\{dG_{n,t}(x_i)/dF(x_i)\} - n^{-1/2} t \sum_{i=1}^n e_F(x_i) + 2^{-1} t^2| > \varepsilon) \to 0 \tag{2.18}$$

as n goes to infinity, for every positive ε, and upon Anderson's (1955) lemma. For an elementary argument, see Beran (1980).

Combining (2.16) with (2.17) and letting k increase to infinity yields Theorem 1. This method of proof is related to Levit and Samarov (1978). Alternate proofs may be based on Millar (1979); the loss function need only be subconvex with respect to the sup-norm.

The lower bound (2.12) is attained asymptotically by estimates $\{\hat{H}_n(x);\ n \geq 1\}$ which have the property described in the next theorem. See (2.9) for the definitions of w_F and h_F.

THEOREM 2. *Suppose that Assumption* 1 *is satisfied and u is bounded. Let* $\{\hat{H}_n(x);\ n \geq 1\}$ *be any sequence of estimates such that*

$$\lim_{n\to\infty}\sup_{n,F,c} P_G(n^{1/2}\|\hat{H}_{n,v} - H_{n,v}(\cdot, F) - \langle h_F, \hat{F}_n - F\rangle w_F\| > \varepsilon) = 0 \tag{2.19}$$

for every positive ε, *every positive c, and every F in* $\mathscr{F}$. *Then*

$$\lim_{n\to\infty}\sup_{n,F,c} R_n(\hat{H}_n, G) = R_0(F) \tag{2.20}$$

for every positive c and every F in $\mathscr{F}$.

PROOF. Because of (2.8), (2.19) is equivalent to the requirement

$$\lim_{n\to\infty}\sup_{n,F,c} P_G(n^{1/2}\|\hat{H}_{n,v} - H_{n,v}(\cdot, G) - \langle h_F, \hat{F}_n - G\rangle w_F\| > \varepsilon) = 0. \tag{2.21}$$

Hence, under every sequence of product measures $\{G_n^n;\ G_n \text{ in } B_n(F, c)\}$, the processes $n^{1/2}\{\hat{H}_{n,v}(x) - H_{n,v}(x, G_n)\}$ converge weakly in sup-norm to the degenerate Gaussian process $a_0 Z w_F(x)$, where Z is a standard normal random variable. (Alternatively, we can say that the processes $n^{1/2}\{\hat{H}_n(x) - H_n(x, G_n)\}$ converge weakly in the norm $\|\cdot\|_v$ to the Gaussian process $a_0 Z x\varphi(x/s(F))$.) This implies Theorem 2.

Two constructions of estimates which satisfy (2.19) or (2.21), and are therefore asymptotically minimax, are as follows.

Bootstrap estimate:

$$\hat{H}_{n,B}(x) = H_n(x, \hat{F}_n). \tag{2.22}$$

This estimate can sometimes be calculated analytically, but is more usually approximated by Monte Carlo methods (Efron, 1979).

First-order Edgeworth expansion estimate:

$$\hat{H}_{n,E}(x) = \Phi(x/s_n(\hat{F}_n)) - n^{-1/2}k(\hat{F}_n)t(x/s_n(\hat{F}_n)) - n^{-1/2}s_n^{-1}(\hat{F}_n)b(\hat{F}_n)\varphi(x/s_n(\hat{F}_n)). \tag{2.23}$$

This estimate requires knowledge of the functionals s_n, k, b and it need not be a distribution function.

To verify the asymptotic behavior of $\{\hat{H}_{n,B}\}$, observe that for every positive c,

$$\lim_{d\to\infty}\limsup_{n\to\infty}\sup_{n,F,c} P_G\{\hat{F}_n \notin B_n(F, d)\} = 0. \tag{2.24}$$

Indeed, for every n, every G in $B_n(F, c)$, and every $d > c$,

$$\begin{aligned} P_G\{\hat{F}_n \notin B_n(F, d)\} &= P_G(n^{1/2}\|\hat{F}_n - F\| > d) \\ &\leq P_G(n^{1/2}\|\hat{F}_n - G\| > d - c) \\ &\leq C\exp\{-(d-c)^2\}, \end{aligned} \tag{2.25}$$

the constant C not depending on G or d (Dvoretzky, Kiefer, and Wolfowitz, 1956). Equation (2.24) is weaker than (2.25) but suffices here; (2.24) and (2.8) imply that $\{\hat{H}_{n,B}\}$ satisfies (2.19).

On the other hand, it follows from (2.7), (2.23), and (2.24) that

$$\lim_{n\to\infty}\sup_{n,F,c} P_G(n^{1/2}\|\hat{H}_{n,B} - \hat{H}_{n,E}\|_v > \varepsilon) = 0 \tag{2.26}$$

for every positive ε and every positive c. Thus, $\{\hat{H}_{n,E}\}$ also satisfies (2.19) and is asymptotically equivalent to the bootstrap estimates.

REMARKS.

(a) Theorems 1 and 2 show that, under Assumption 1, both the bootstrap estimate $\hat{H}_{n,B}$ and the first-order Edgeworth expansion estimate $\hat{H}_{n,E}$ are at least as good as *any* other estimate of $H_n(x, F)$, in an asymptotically minimax sense.

(b) This asymptotic minimax property may also be interpreted as quantitative robustness over the contamination neighborhood $B_n(F, c)$ (cf. Beran, 1981).

(c) Let F_n^* be an estimate of F other than empirical d.f. When is the resampling estimate $H_n(x, F_n^*)$ asymptotically minimax? Answer: whenever F_n^* satisfies (2.24), and the limiting distribution of $\{n^{1/2}\langle h_F, F_n^* - G_n\rangle\}$ under every sequence $\{G_n^n; G_n \text{ in } B_n(F, c)\}$ is $N(0, a_0^2)$; for then the risk of $H_n(x, F_n^*)$ satisfies (2.20). A simple sufficient condition is that

$$\lim_{n\to\infty}\sup_{n,F,c} P_G(n^{1/2}\|F_n^* - \hat{F}_n\| > \varepsilon) = 0 \tag{2.27}$$

for every positive ε, every positive c, and every F in $\mathscr{F}$.

(d) The most commonly used estimate of $H_n(x, F)$, the normal approximation $\Phi(x/s_n(\hat{F}_n))$, is not, in general, asymptotically minimax. Indeed, Assumption 1 and equations (2.7), (2.8) imply

$$\lim_{n\to\infty}\sup_{n,F,c}\| n^{1/2}\{\Phi_v(x/s_n(G)) - H_{n,v}(x, F)\} - n^{1/2}\langle h_F, G - F\rangle w_F(x) - d_F(x)\| = 0, \tag{2.28}$$

where

$$d_F(x) = k(F)t_v(x/s(F)) + s^{-1}(F)b(F)\varphi_v(x/s(F)). \tag{2.29}$$

It follows from (2.28), (2.8), and (2.24) that

$$\lim_{n\to\infty}\sup_{n,F,c} P_G(n^{1/2}\|\Phi_v(\cdot/s_n(\hat{F}_n)) - H_{n,v}(\cdot, G) - \langle h_F, F_n - G\rangle w_F - d_F\| > \varepsilon) = 0 \tag{2.30}$$

for every positive ε, every positive c, and every F in $\mathscr{F}$. Thus, under every sequence of product measures $\{G_n^n; G_n \text{ in } B_n(F, c)\}$, the processes $n^{1/2}\{\Phi_v(x/s_n(\hat{F}_n)) - H_{n,v}(x, G_n)\}$ converge weakly in sup-norm to the degenerate Gaussian process $a_0 Z w_F(x) + d_F(x)$, where Z is a standard normal random variable. In other words, the processes $n^{1/2}\{\Phi(x/s_n(\hat{F}_n)) - H_n(x, G_n)\}$ converge weakly in the norm $\|\cdot\|_v$ to the Gaussian process $a_0 Z x\varphi(x/s(F)) + d_F^\#(x)$; the function $d_F^\#$ is defined by dropping the subscript v in (2.29).

From these calculations it follows that

$$\lim_{n\to\infty}\sup_{n,F,c} R_n(\Phi(\cdot/s_n(\hat{F}_n)), G) = Eu(\|a_0 Z w_F + d_F\|). \tag{2.31}$$

By Theorem 1, the right side of (2.31) is not less than $R_0(F)$. Equality occurs when $d_F(x)$ vanishes; that is, when F is such that $k(F) = 0$ and the centering functional $T_n(F)$ matches $m_n(F)$ up to terms of order n^{-1}, so that $b(F) = 0$.

(e) As an estimate of $H_n(x, F)$, the normal approximation $\Phi(x/s_n(\hat{F}_n))$ has the right rate of convergence and even the right asymptotic covariance function, but is biased. The asymptotically minimax estimate $\hat{H}_{n,E}$ is simply $\Phi(x/s_n(\hat{F}_n))$ minus the natural estimate of the bias.

2.2 *Estimating functionals of the sampling distribution* $H_n(x, F)$. Also of interest are certain functionals of $H_n(x, F)$, particularly location and scale functionals such as the mean and standard deviation or their more robust competitors. The mean of $H_n(x, F)$ is the bias of $\hat{T}_n$ as an estimate of $T_n(F)$; the variance of $H_n(x, F)$ is the variance of $\hat{T}_n$.

Let $\{U_n; n \geq 1\}$ be a sequence of real-valued functionals of distribution functions on the real line. The problem is to estimate $V_n(F) = U_n(H_n(\cdot, F))$, which can be regarded either as a functional of $H_n(x, F)$ or as a functional of F. Both points of view prove useful in what follows. As competing estimates of $V_n(F)$, we will consider all functions of the sample which can be represented in the form $\hat{V}_n = U_n(\hat{H}_n)$, where $\hat{H}_n$ is any function mapping the sample into a distribution function on the real line. The technical reasons for working with this particular large class of estimates will become clearer in the next paragraph.

The risk of $\hat{V}_n = U_n(\hat{H}_n)$ as an estimate of $V_n(G) = U_n(H_n(\cdot, G))$ is defined to be

$$r_n(\hat{H}_n, G) = E_G u(n^{1/2} | \hat{V}_{n,v} - V_{n,v}(G)|) \tag{2.32}$$

where $u: R^+ \to R^+$ is monotone increasing and

$$\hat{V}_{n,v} = U_n(\hat{H}_{n,v}), \qquad V_{n,v}(G) = U_n(H_{n,v}(\cdot, G)) \tag{2.33}$$

for $\hat{H}_{n,v}$, $H_{n,v}(\cdot, G)$ defined as in Section 2.1. This risk, a perturbation of the more familiar $E_G u(n^{1/2} | \hat{V}_n - V_n(G)|)$, is introduced to avoid difficulties with the lattice distributions G in $B_n(F, c)$.

Let ν be a sigma-finite measure on the real line and let $\langle \cdot, \cdot \rangle_1$ denote the inner product with respect to ν.

ASSUMPTION 2. The functionals $\{U_n\}$ are differentiable at $\{\Phi_v(x/s_n(F))\}$ in the following sense: there exists u_F in $L^1(\nu)$ such that

$$\lim_{n\to\infty} \sup_H n^{1/2} | U_n(H) - U_n(\Phi_v(\cdot/s_n(F))) - \langle u_F, H - \Phi_v(\cdot/s_n(F))\rangle_1 | = 0 \tag{2.34}$$

for every positive d and every F in $\mathscr{F}$. The supremum is taken over all distribution functions H in $B_n(\Phi_v(\cdot/s_n(F)), d)$.

The clipped mean and variance

$$U_1(H) = \int_{-B}^{B} x\, dH(x), \qquad U_2(H) = \int_{-B}^{B} x^2\, dH(x) - U_1^2(H) \tag{2.35}$$

both satisfy Assumption 2. Taking ν to be Lebesgue measure on the interval $(-B, B)$ plus unit masses at B and $-B$, the derivatives are

$$\begin{aligned} u_{1,F}(x) &= -1 + BI_{\{-B,B\}}(x), \\ u_{2,F}(x) &= -2\{x - U_1(H)\} + \{B^2 + 2BU_1(H)\}I_{\{-B,B\}}(x) \end{aligned} \tag{2.36}$$

respectively. A variety of other M and L functionals satisfy Assumption 2 (cf. Boos, 1979; Boos and Serfling, 1980).

Equation (2.7) and Assumption 1 imply the following: if G lies in $B_n(F, c)$, there exists positive d which does not depend on G such that $H_{n,v}(x, G)$ lies in $B_n(\Phi_v(\cdot/s_n(F)), d)$ for all $n \geq n_0(F)$. Hence, by Assumption 2,

$$\begin{aligned} \lim_{n\to\infty} \sup_{n,F,c} n^{1/2} | V_{n,v}(G) &- U_n(\Phi_v(\cdot/s_n(F))) \\ &- \langle u_F, H_{n,v}(\cdot, G) - \Phi_v(\cdot/s_n(F))\rangle_1 = 0. \end{aligned} \tag{2.37}$$

Combining (2.37) with (2.8) yeilds

$$\lim_{n\to\infty} \sup_{n,F,c} n^{1/2} | V_{n,v}(G) - V_{n,v}(F) - \langle v_F, G - F\rangle | = 0, \tag{2.38}$$

where

$$v_F(x) = \langle u_F, w_F\rangle_1 h_F(x) \tag{2.39}$$

for h_F and w_F defined in (2.9).

The following theorem is a consequence of (2.38) and Hájek (1972), analogous to Theorem 1. See Koshevnik and Levit (1975) for a proof.

THEOREM 3. *Suppose Assumptions* 1 *and* 2 *are satisfied. Then, for every F in* $\mathscr{F}$,

$$\lim_{c\to\infty}\lim\inf_{n\to\infty}\inf_{\hat{H}_n}\sup_{n,F,c} r_n(\hat{H}_n, G) \geq r_0(F), \tag{2.40}$$

where

$$r_0(F) = \int u(b\,|\,z\,|)\varphi(z)\,dz \tag{2.41}$$

and

$$b = \left\{\int v_{F,0}^2(x)\,dF(x)\right\}^{1/2}. \tag{2.42}$$

The function $v_{F,0}$ *is obtained by replacing* h_F *in* (2.11) *with* v_F.

The asymptotic minimax bound $r_0(F)$ in (2.40) is attained by the natural bootstrap estimate of $V_n(F) = U_n(H_n(\cdot, F))$:

$$\hat{V}_{n,B} = U_n(H_n(\cdot, \hat{F}_n)). \tag{2.43}$$

To verify this assertion, observe that in the notation of (2.33),

$$\hat{V}_{nB,v} = U_n(H_{n,v}(\cdot, \hat{F}_n)) = V_{n,v}(\hat{F}_n). \tag{2.44}$$

It follows from (2.24) and (2.38) that for every positive ε, every positive c, and every F in $\mathscr{F}$,

$$\lim_{n\to\infty}\sup_{n,F,c} P_G(n^{1/2}\,|\,\hat{V}_{n,B,v} - V_{n,v}(F) - \langle v_F, \hat{F}_n - F\rangle\,| > \varepsilon) = 0. \tag{2.45}$$

Equivalently, using (2.38) again,

$$\lim_{n\to\infty}\sup_{n,F,c} P_G(n^{1/2}\,|\,\hat{V}_{n,B,v} - V_{n,v}(G) - \langle v_F, \hat{F}_n - G\rangle\,| > \varepsilon) = 0. \tag{2.46}$$

Equation (2.46) implies that

$$\lim_{n\to\infty}\sup_{n,F,c} r_n(\hat{H}_{n,B}, G) = r_0(F). \tag{2.47}$$

Indeed, under every sequence of product measures $\{G_n^n;\ G_n \text{ in } B_n(F, c)\}$, the centered estimates $n^{1/2}\{\hat{V}_{n,B,v} - V_{n,v}(G_n)\}$ converge weakly to bZ, where Z is a standard normal random variable.

Theorem 3 and (2.47) show that the bootstrap estimate $\hat{V}_{n,B}$ for $V_n(F) = U_n(H(\cdot, F))$ cannot be beaten, in an asymptotically minimax sense, by any other estimate of the form $U_n(\hat{H}_n)$.

2.3 *Estimating center-of-symmetry functionals of* $H_n(x, F)$. The discussion in Section 2.2 becomes trivial if v_F, defined in (2.39), vanishes. Since w_F is an odd function, v_F will vanish whenever the derivative u_F of $\{U_n\}$ is even and the measure ν is symmetric about the origin. The clipped mean defined in (2.35) is an example of such a functional. Indeed, the vanishing of v_F is to be expected for any differentiable location functional whose value at every symmetric distribution is the center-of-symmetry. More interesting optimality results for this case can be obtained by replacing the $n^{-1/2}$ rate of convergence in Section 2.2 with an n^{-1} rate and by modifying Assumptions 1 and 2 accordingly.

ASSUMPTION 1′. Let $J_n(x, G)$ be the distribution function of $n^{1/2}(\hat{T}_n - m_n(G))/s_n(G)$ under G. The following assertions hold for every positive c and every F in $\mathscr{F}$:

(a) $J_{n,v}(x, G)$ has a second order Edgeworth expansion

$$\begin{aligned}\lim_{n\to\infty}\sup_{n,F,c}\|\, n\{J_{n,v}(x, G) - \Phi_v(x)\} &+ n^{1/2}k_3(G)t_{1,v}(x)\\ &+ k_4(G)t_{2,v}(x) + k_3^2(G)t_{3,v}(x)\,\| = 0,\end{aligned} \tag{2.48}$$

where

$$t_1(x) = 6^{-1}(x^2 - 1)\varphi(x), \qquad t_2(x) = 24^{-1}(x^3 - 3x)\varphi(x),$$
$$t_3(x) = 72^{-1}(x^5 - 10x^3 + 15x)\varphi(x) \tag{2.49}$$

and $k_3(G)$, $k_4(G)$ are functionals of G. The subscript v indicates convolution with v.

(b) There exist functionals $b_1(G)$, $b_2(G)$ such that

$$\lim_{n\to\infty}\sup_{n,F,c} n^{3/2}\,|\,m_n(G) - T_n(G) - n^{-1}b_1(G) - n^{-3/2}b_2(G)\,| = 0. \tag{2.50}$$

(c) The functionals $\{s_n(G)\}$, $b_1(G)$, and $k_3(G)$ are differentiable at F in the sense (2.5), with respective derivatives s'_F, $b'_{1,F}$, $k'_{3,F}$ in $L^1(\mu)$. Moreover, $s(F) = \lim_{n\to\infty} s_n(F)$ exists.

(d) The functionals $b_2(G)$ and $k_4(G)$ are $\|\cdot\|$ continuous at F.

Suppose $\hat{T}_n$ is the sample mean and $\mathscr{F}$ is the set of all distribution functions supported on a fixed compact interval. A weaker form of Assumption 1′, in which $B_n(F, c)$ is replaced by $B_n(F, c) \cap \mathscr{F}$, holds for $\hat{T}_n$. Refer to the discussion following Assumption 1 in Section 2.1, noting that in this instance $k_r(G)$ is the rth cumulant of G divided by the rth power of the standard deviation of G. Circumstances under which U-statistics satisfy Assumption 1′ unmodified are described in Section 3.

Calculations using (2.6) and Assumption 1′ yield

$$\begin{aligned}\lim_{n\to\infty}\sup_{n,F,c} \| \, & n\{H_{n,v}(x, G) - \Phi_v(x/s_n(G))\} \\ & + n^{1/2}\{c_{1n}(G)\varphi_v(x/s_n(G)) + k_3(G)t_{1,v}(x/s_n(G))\} \\ & + c_{2n}(G)\varphi_v(x/s_n(G)) - 2^{-1}c_{1n}^2(G)\varphi'_v(x/s_n(G)) \\ & - k_3(G)c_{1n}(G)t'_{1,v}(x/s_n(G)) + k_4(G)t_{2,v}(x/s_n(G)) \\ & + k_3^2(G)t_{3,v}(x/s_n(G))\,\| = 0,\end{aligned} \tag{2.51}$$

where $c_{in}(G) = s_n^{-1}(G)b_i(G)$. Also, by Taylor expansion,

$$\begin{aligned}\lim_{n\to\infty}\sup_{n,F,c}\, n \, \| \, & \Phi_v(x/s_n(G)) - \Phi_v(x/s_n(F)) \\ & + s_n^{-2}(F)(s_n(G) - s_n(F))x\varphi_v(x/s_n(F)) \\ & - 2^{-1}\{s_n(G) - s_n(F)\}^2 \\ & \cdot\{s_n^{-4}(F)x^2\varphi'_v(x/s_n(F)) + 2s_n^{-3}(F)x\varphi_v(x/s_n(F))\}\,\| = 0.\end{aligned} \tag{2.52}$$

The essential behavior of twice differentiable center-of-symmetry functionals is captured in the following assumption.

ASSUMPTION 2′. The functionals $\{U_n\}$ are twice differentiable at $\{\Phi_v(x/s_n(F))\}$ in the following sense: there exist u_F in $L^1(\nu)$ and q_F in $L^1(\nu) \times L^1(\nu)$ such that, for every positive d and every F in $\mathscr{F}$,

$$\begin{aligned}\lim_{n\to\infty}\sup_H n\,|\,& U_n(H) - U_n(\Phi_v(\cdot/s_n(F))) - \langle u_F, H - \Phi_v(\cdot/s_n(F))\rangle_1 \\ & - \langle q_F(H - \Phi_v(\cdot/s_n(F))), H - \Phi_v(\cdot/s_n(F))\rangle_1\,| = 0,\end{aligned} \tag{2.53}$$

where $q_F h(x) = \int q_F(x, y)h(y)\,d\nu(y)$. The supremum is taken over all distribution functions H in $B_n(\Phi_v(\cdot/s_n(F)), d)$. Moreover, $\langle u_F, h\rangle_1 = \langle q_F h, h\rangle_1 = 0$ for every bounded odd function h.

The clipped mean satisfies Assumption 2′ with ν and u_F as described in (2.36) and with q_F identically zero. Other M and L functionals for center-of-symmetry yield nontrivial q_F.

As in Section 2.2, let $V_n(G) \equiv U_n(H_n(\cdot, G))$ denote the functional to be estimated. Calculations based upon (2.51), (2.52) and Assumptions 1′ and 2′ eventually yield

$$\lim_{n\to\infty}\sup_{n,F,c}|\,n\{V_{n,v}(G) - V_{n,v}(F)\} - n^{1/2}\langle v_F^*, G - F\rangle\,| = 0, \tag{2.54}$$

where

$$v_F^{\#} = \langle u_F, \varphi_v(x/s(F))\rangle_1 c'_{1,F} + \langle u_F, t_{1,v}(x/s(F))\rangle_1 k'_{3,F} - \langle u_F, c_1(F)x\varphi'_v(x/s(F)) + k_3(F)xt'_{1,v}(x/s(F))\rangle_1 h_F \tag{2.55}$$

and

$$c_1(F) = s^{-1}(F)b_1(F), \qquad c'_{1,F} = s^{-2}(F)\circ\{s(F)b'_{1,F} - b_1(F)s'_F\}. \tag{2.56}$$

Let $\hat{V}_n$ be any estimate for $V_n(G)$ of the form $U_n(\hat{H}_n)$, where $\hat{H}_n$ is an estimate of $H_n(x, G)$. It is clear, comparing (2.54) with (2.38), that an analog of Theorem 3 holds for the functionals considered in this section. If $n^{1/2}$ is changed to n in the risk (2.32), then v_F in Theorem 3 is replaced by $v_F^{\#}$. Moreover, the bootstrap estimate $\hat{V}_{n,B} = U_n(\hat{H}_{n,B})$ is still an asymptotically minimax estimate of $V_n(G)$; and the asymptotic distribution of $n\{\hat{V}_{n,B,v} - V_{n,v}(G_n)\}$ is normal with mean zero under every sequence of product measures $\{G_n^n; G_n$ in $B_n(F, c)\}$.

3. Edgeworth expansions. It is not immediately obvious that the locally uniform Edgeworth expansions postulated in Assumptions 1 and 1′ actually exist. The purpose of this section is to identify a useful group of statistics to which the theory of this paper is applicable. When $\hat{T}_n$ is the sample mean or a function of the sample mean, Singh (1981) and Bickel and Freedman (1980) have obtained Edgeworth expansions for $J_n(x, \hat{F}_n)$. Their results are related to the discussion here since equation (2.24) and part (a) of Assumptions 1 or 1′ imply Edgeworth expansions for the smoothed $J_{n,v}(x, \hat{F}_n)$.

3.1 *U-statistics.* Consider the second degree U-statistic

$$\hat{T}_n = 2n^{-1}(n-1)^{-1} \textstyle\sum_{i<j} t(X_i, X_j) \tag{3.1}$$

where $t(x, y)$ is symmetric in its arguments. Assume that t is absolutely continuous, vanishes outside a large square $[-B, B]^2$, and has essentially bounded derivative. The assumption that t ultimately vanishes is harmless, practically speaking, provided the square is chosen large enough to contain the domain of computation.

Define $m(G) = E_G t(X_1, X_2)$. Then

$$n^{1/2}\{\hat{T}_n - m(G)\} = 2n^{-1/2}(n-1)^{-1} \textstyle\sum_{i<j} h_G(X_i, X_j) \tag{3.2}$$

for $h_G(X_i, X_j) = t(X_i, X_j) - m(G)$. Put

$$g_G(X_i) = E_G\{h(X_i, X_j) \mid X_i\}, \qquad d_G(X_i, X_j) = h_G(X_i, X_j) - g_G(X_i) - g_G(X_j) \tag{3.3}$$

and define

$$\begin{aligned} s_n^2(G) &= 4s_G^2 + 2(n-1)^{-1}E_G\, d_G^2(X_1, X_2) \\ k_3(G) &= s_G^{-3}[E_G g_G^3(X_1) + 3E_G\{g_G(X_1)g_G(X_2)\, d_G(X_1, X_2)\}] \\ k_4(G) &= s_G^{-4}[E_G g_G^4(X_1) - 3s_G^4 + 12E_G\{g_G^2(X_1)g_G(X_2)\, d_G(X_1, X_2)\} \\ &\qquad + 12E_G\{g_G(X_2)g_G(X_3)\, d_G(X_1, X_2)\, d_G(X_1, X_3)\}], \end{aligned} \tag{3.4}$$

where $s_G^2 = E_G g_G^2(X_1)$.

If μ is the restriction of Lebesgue measure to the interval $[-B, B]$, the functionals $m(G)$, $\{s_n(G)\}$, $k_3(G)$, and $k_4(G)$ are each differentiable in the sense (2.5). Indeed, let $t_{11}(x, y)$ be the derivative of $t(x, y)$, so that

$$t(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{x} t_{11}(u, v)\, du\, dv \tag{3.5}$$

for every x, y. Let

$$t_{0,1}(x, v) = \int_{-\infty}^{x} t_{11}(u, v)\, du. \tag{3.6}$$

The assumptions on $t(x, y)$ imply that $t_{11}(u, v)$ vanishes a.e. outside $[-B, B]^2$ and that $t_{0,1}(x, v) = 0$ for almost every v whenever x is outside $[-B, B]$. Integration by parts yields

$$m(G) = \int G(x)G(y)t_{11}(x, y)\, dx\, dy \tag{3.7}$$

and

$$g_G(x) = -\int G(y)t_{0,1}(x, y)\, dy - m(G). \tag{3.8}$$

The sup-norm differentiability of $m(G)$ is immediate. Using (3.8), the decomposition

$$s_G^2 - s_F^2 = \int g_G^2\, d(G - F) + \int (g_G - g_F)(g_G + g_F)\, dF, \tag{3.9}$$

and integration by parts in the first term on the right side of (3.9) establishes the differentiability of s_G^2. By continuing in this fashion, we verify the differentiability of $s_n^2(F)$ (and hence of $s_n(F)$), $k_3(G)$, and $k_4(G)$.

Let

$$\tilde{J}_n(x, G) = \Phi(x) - n^{-1/2}k_3(G)t_1(x) - n^{-1}k_4(G)t_2(x) - n^{-1}k_3^2(G)t_3(x). \tag{3.10}$$

To establish the locally uniform second order Edgeworth expansion (2.48), it suffices to show that for every sequence $\{G_n \text{ in } B_n(F, c)\}$,

$$\lim_{n\to\infty}\sup_{n,F,c} n \,\| J_{n,v}(x, G_n) - \tilde{J}_{n,v}(x, G_n) \| = 0, \tag{3.11}$$

whatever the choice of positive c and F in $\mathscr{F}$.

Let v be the smoothing density (2.1). The characteristic function of v is

$$\psi(t) = 2(at)^{-2}\{1 - \cos(at)\}. \tag{3.12}$$

Let $j_n(t)$, $\tilde{j}_n(t)$ be the characteristic functions of $J_n(x, G_n)$, $\tilde{J}_n(x, G_n)$ respectively. Similarly, let $j_{n,v} = j_n\psi$ and $\tilde{j}_{n,v} = \tilde{j}_n\psi$ denote the characteristic functions of the smoothed $J_{n,v}(x, G_n)$ and $\tilde{J}_{n,v}(x, G_n)$. By Esséen's lemma (see Feller, 1966, page 512 and Callaert et al., 1980, page 301),

$$\| J_{n,v}(x, G_n) - \tilde{J}_{n,v}(x, G_n) \| \le 2\pi^{-1} \int_0^{n \log n} t^{-1} | j_{n,v}(t) - \tilde{j}_{n,v}(t) |\, dt + o(n^{-1}). \tag{3.13}$$

The integral on the right side of (3.13) can be subdivided into three integrals, whose ranges of integration are 0 to $n^{1/4}/\log n$, $n^{1/4}/\log n$ to $n^{3/4}/\log n$, and $n^{3/4}/\log n$ to $n \log n$. Label these integrals I, II, III respectively. Since $| \psi(t) | \le 4(at)^{-2}$, integral III is $o(n^{-1})$.

That integrals I and II are also $o(n^{-1})$ - a fact which would imply (3.11) - can be checked by examining the reasoning in Sections 3 and 4 of Callaert et al. (1980). Their fixed F argument works here for every sequence $\{G_n \text{ in } B_n(F, c)\}$ for the following reasons.

(a) Moments calculated under G_n of polynomials in $g_G(X_i)$ and $d_G(X_j, X_k)$ converge to the corresponding moments calculated under F. This follows from our strong assumptions on $t(x, y)$. The various bounds in their argument which involve such moments remain asymptotically valid under $\{G_n\}$.

(b) The fact $j_{n,v} = j_n\psi$ and the rapid decay of $| \psi(t) |$ as $| t |$ increases simplifies the treatment of integral II in their Section 4.

Their conditions B and C are unnecessary in analyzing integrals I, II, III above because of the smoothing.

The sample mean and sample variance are U-statistics corresponding to $t(x, y) = 2^{-1}(x + y)$ and $t(x, y) = 2^{-1}(x - y)^2$ respectively. These functions $t(x, y)$ do not satisfy the assumptions stated at the beginning of this section. One practical resolution is to assume that the distributions G are all supported on an interval $[-C, C]$, where $C < B$. If so, $t(x, y)$ can be modified arbitrarily outside $[-C, C]^2$ without affecting either the estimate $\hat{T}_n$ or the associated functionals $m(G)$, $s_n(G)$, $k_r(G)$.

3.2 *Second order von Mises expansions.* Let $\{\hat{T}_n\}$ be statistics and let $\{T_n(F)\}$ be centering functionals such that

$$\lim_{n\to\infty}\sup_{n,F,c} n^2 E_G\{\hat{T}_n - T_n(F) - n^{-2}\textstyle\sum_{i,j} t_F(X_i, X_j)\}^2 = 0, \tag{3.14}$$

where $t_F(x, y)$ is symmetric in its arguments. This class of statistics contains M-estimates, L-estimates, and certain maximum likelihood estimates (cf. Serfling, 1980, Chapters 6 to 8). Equation (3.14) is a locally uniform second order von Mises expansion of T_n. Assume that $t_F(x, y)$ is absolutely continuous, vanishes outside a large square $[-B, B]^2$, and has essentially bounded derivative. Rearrangement of (3.14) then yields

$$\lim_{n\to\infty}\sup_{n,F,c} n^2 E_G\{\hat{T}_n - m_n(G) - 2n^{-1}(n-1)^{-1}\textstyle\sum_{i<j} h_G(X_i, X_j)\}^2 = 0, \tag{3.15}$$

for

$$\begin{aligned} h_G(X_i, X_j) &= t_F(X_i, X_j) - E_G t_F(X_1, X_2) \\ m_n(G) &= T_n(F) + E_G t_F(X_1, X_2) + n^{-1}\{E_G t_F(X_1, X_1) - E_G t_F(X_1, X_2)\}. \end{aligned} \tag{3.16}$$

Approximation (3.15) suggests that $J_{n,v}(x, G)$ has the first order Edgeworth expansion (2.3) with $m_n(G)$ as above and with $s_n^2(G)$, $k(G) = k_3(G)$ as in (3.4). The measure μ is again Lebesgue measure restricted to the interval $[-B, B]$. To justify these claims, let $J^*_{n,v}(x, G)$ denote the smoothed sampling distribution of $2n^{-1/2}(n-1)^{-1}\sum_{i<j} h_G(X_i, X_j)$ under G. Let $j_{n,v}(t)$ and $j^*_{n,v}(t)$ be the characteristic functions of $J_{n,v}(x, G)$ and $J^*_{n,v}(x, G)$ respectively. Set

$$Z_n = n\{\hat{T}_n - m_n(G) - 2n^{-1}(n-1)^{-1}\textstyle\sum_{i<j} h_G(X_i, X_j)\}. \tag{3.17}$$

Evidently,

$$|j^*_{n,v}(t) - j_{n,v}(t)| \le E_G|\exp(itn^{-1/2}Z_n) - 1| \cdot |\psi(t)|. \tag{3.18}$$

Taylor expansion of the exponential term, Esséen's lemma, and (3.15) yield

$$\lim_{n\to\infty}\sup_{n,F,c} n^{1/2}\,\|J_{n,v}(\cdot, G) - J^*_{n,v}(\cdot, G)\| = 0. \tag{3.19}$$

In view of Section 3.1, this implies the desired first order Edgeworth expansion for $J_{n,v}(x, G)$.

Acknowledgments. I am grateful to Brad Efron for comments on the Hodges estimate example and to the referees for expository suggestions.

REFERENCES

ANDERSON, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6** 170–176.

BERAN, R. (1980). Asymptotic lower bounds for risk in robust estimation. *Ann. Statist.* **8** 1252–1264.

BERAN, R. (1981). Efficient robust estimates in parametric models. *Z. Wahrsch. verw. Gebiete* **55** 91–108.

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

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

BOOS, D. D. (1979). A differential for L-statistics. *Ann. Statist.* **7** 955–959.

BOOS, D. D. and SERFLING, R. J. (1980). A note on differentials and the CLT and LIL for statistical functions with applications to M-estimates. *Ann. Statist.* **8** 618–624.

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

CALLAERT, H., JANSSEN, P., and VERAVERBEKE, N. (1980). An Edgeworth expansion for *U*-statistics. *Ann. Statist.* **8** 299–312.

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.

FELLER, W. (1966). *An Introduction to Probability Theory and its Applications* **2.** Wiley, New York.

HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 175–194. Univ. of California Press.

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

LECAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. of California Publications in Statist.* **1** 277–330.

LEVIT, B. YA and SAMAROV, A. M. (1978). Estimation of spectral functions. *Problems Inform. Transmission* **14** 120–124.

MILLAR, P. W. (1979). Asymptotic minimax theorems for the sample distribution function. *Z. Wahrsch. verw. Gebiete* **48** 233–252.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

SINGH, K. (1981). On asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CA 94720