The Annals of Statistics
1988, Vol. 16, No. 1, 450–469

FUNCTIONAL JACKKNIFING: RATIONALITY AND GENERAL ASYMPTOTICS[1]

BY PRANAB KUMAR SEN

University of North Carolina, Chapel Hill

Though jackknifing serves well the dual purpose of bias reduction and variance estimation, the pseudovariables it generates may not generally preserve robustness for general statistical functionals. These pseudovariables are incorporated in a differentiable functional detour of jackknifing, and along with its rationality, the related asymptotic theory is studied systematically. A two-step jackknifing is considered for the variance estimation. Second-order asymptotic distributional representations for the classical jackknifed estimators are also considered.

1. Introduction. The *jackknifing*, originally conceived for possible reduction of (smaller-order) bias of an estimator, generates *pseudovariables*, which provide a (strongly) consistent estimator of the (asymptotic) *variance* of the estimator (as well as its jackknifed version). For some detailed studies of this dual role of jackknifing, refer to Miller (1974), Sen (1977) and Parr (1985), among others. To filter *robustness* under jackknifing, one should start with a robust initial estimator; otherwise, the pseudovariables (generated by jackknifing) may lead to a less robust jackknifed version. By their very construction, the pseudovariables are more vulnerable to outliers and error contaminations, and hence the classical jackknifed estimator (being their simple arithmetic mean) may not be the ideal choice for a robust and adaptive estimator.

For a general *statistical functional* $\theta = T(F)$, for an estimator $\hat{\theta}_n$ based on n independent and identically distributed random variables (i.i.d.r.v.'s) $X_1, \dots, X_n$, drawn from the distribution F, the lack of robustness of the classical jackknifed estimator (say, θ_n^*) has been noticed by many workers. For this reason, Hinkley and Wang (1980) and Parr (1985) considered alternative ways of recombining the pseudovariables for a more robust version. The first paper deals with trimmed jackknifing, whereas Parr's suggestion relates to a variant of the L-functional approach. Incidentally, Parr's definition of this L-functional jackknifing may run into difficulties when $\hat{\theta}_n$ or θ_n^* is real-valued but the X_i are vector-valued (viz., the sample correlation coefficient). The main strength of Parr's paper, of course, lies in its unified treatment of jackknifed estimators of general statistical functionals under a *second-order* (strong) Fréchet differentiability condition. To illustrate this point, he has elaborated on the validity of the first-order strong Fréchet differentiability for some important types of statistical functionals, and presumably, this analysis may as well be extended to the second-order case. But

Received October 1986; revised June 1987.
[1]Work supported by Office of Naval Research Contract N00014-83-K-0387.
AMS 1980 *subject classifications*. 60F05, 62E20, 62F12.
Key words and phrases. Asymptotic normality, asymptotic variance, bias, bootstrap, Fréchet differentiability, Hadamard derivative, jackknifed variance, perturbations, pseudovariables, robustness, second-order representation, statistical functional, two-step jackknifing.

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

the end product may then look less appealing. Thus there remains some scope for reexamining these regularity conditions, even in a broader perspective of incorporating a more general functional of the pseudovariables in the formulation of jackknifed estimators.

Note that both the works of Hinkley and Wang (1980) and Parr (1985) relate to some *trimming* on the pseudovariables to enhance robustness. This may as well be achieved by incorporating other forms of functionals of these pseudovariables without this explicit trimming. The main objective of the current study is to focus on such a general class of functional jackknifing and present the related asymptotic theory in a systematic manner. This class contains the classical and trimmed jackknifing as special cases. [Technically, Parr's (1985) estimator may not generally belong to this class, although a variant form of it does.] For some simple functionals, the close relationship between *bootstrapping* and jackknifing has been studied by various workers [viz., Efron (1982) and Parr (1983), among others]. A similar picture holds for other functionals too. Thus it seems quite plausible to propose a functional bootstrapping, although from robustness considerations, bootstrapping is better than jackknifing (as it does not involve the pseudovariables), and hence we may have a lesser need for functional bootstrapping. Therefore we refrain from the study of functional bootstrapping.

One of the main attractions of the classical jackknifing is that the *jackknifed variance estimator* is strongly consistent for the asymptotic variance of the original estimator (as well as its jackknifed version). However, it has been observed by Hinkley and Wang (1980) and Parr (1985) that any detour from the classical jackknifing may distort this feature. Hinkley and Wang (1980) considered some variance estimators for their trimmed jackknifing, whereas (in view of the fact that the Fréchet derivative depends on the unknown F), Parr's treatment remains a bit incomplete in this respect. In the context of general functional jackknifing, there is thus a genuine need to provide a variance estimator that not only can be used to draw statistical inference but also can be used to assess its (asymptotic) efficiency relative to the classical jackknifing. Generally, to enhance robustness through functional jackknifing, one may entail minor loss of efficiency relative to the classical jackknifing, and a representative picture of this relative loss can be drawn from the respective variance estimators. For this reason, variance estimation in functional jackknifing is also considered here. In this context, a *two-stage jackknifing procedure* is considered, which serves this purpose effectively.

It turns out that for a general statistical functional, the classical jackknifing relates essentially to an adjustment for bias of the original estimator [viz., Parr (1985)]. Under the same regularity conditions, a *second-order asymptotic distributional representation* (SOADR) for the classical jackknifing is considered. This result differs from the parallel SOADR results for M-estimators of location, studied by Jurečková (1985) and Jurečková and Sen (1987), among others.

We find it more convenient to work with the *Hadamard* (*or compact*) *differentiability* of statistical functionals [than the strong Fréchet differentiability, treated in Parr (1985) and other places]. In this context, the *Hadamard continuity* of statistical functionals has also been used. For some general equivalence results, refer to Fernholz (1983) and Parr (1985), although Fernholz's treatment signals a clear superiority for the Hadamard case. This point will be elaborated further in subsequent sections.

Along with the preliminary notions, the SOADR results for the classical jackknifing are considered in Section 2. Section 3 deals with the general formulation and rationality of functional jackknifing. Allied asymptotic theory is presented in Section 4. Variance estimation in functional jackknifing is considered in Section 5. Some general remarks are made in the concluding section.

2. Preliminary notions and SOADR results. Let $\mathscr{L}(A, B)$ be the set of continuous linear transformations from a topological vector space A to another B, and let $\mathscr{C}$ be a class of compact subsets of A, such that every subset consisting of a single point belongs to $\mathscr{C}$. Also, let A^0 be an open subset of A. A function T: $A^0 \to B$ is said to be *Hadamard* (or *compact*) *differentiable* at $F \in A$, if there exists a $T_F' \in \mathscr{L}(A, B)$, such that for any $K \in \mathscr{C}$,

$$\lim_{t \to 0} \{t^{-1}[T(F + tJ) - T(F) - T_F'(tJ)]\} = 0, \tag{2.1}$$

uniformly for $J \in K$; T_F' is called the *compact derivative* of T at F. In the context of jackknifing, usually, we need the *second-order compact differentiability* of T (at F), that is, we assume that for any $K \in \mathscr{C}$,

$$\begin{aligned} T(G) = T(F + (G - F)) = T(F) &+ \int T_1(F; x)\, d[G(x) - F(x)] \\ &+ \tfrac{1}{2} \iint T_2(F; x, y)\, d[G(x) - F(x)] \\ &\times d[G(y) - F(y)] + R_2(F; G - F), \end{aligned} \tag{2.2}$$

where

$$|R_2(F; G - F)| = o(\|G - F\|^2), \quad \text{uniformly in } G \in K, \tag{2.3}$$

and $\|G - F\|$ refers to the usual *sup-norm* [i.e., $\sup_x |G(x) - F(x)|$]. The functions $T_1(F; \cdot)$ and $T_2(F; \cdot)$ are called the *first-* and *second-order compact derivatives* of $T(\cdot)$ (at F), and we can always normalize them in such a way that

$$\int T_1(F; x)\, dF(x) = 0, \qquad T_2(F; x, y) \equiv T_2(F; x, y), \tag{2.4}$$

$$\int T_2(F; x, y)\, dF(y) \equiv 0 \equiv \int T_2(F; y, x)\, dF(x), \tag{2.5}$$

Consider also the functional

$$T_2^*(G) = \int T_2(G; x, x)\, dG(x), \qquad G \in A. \tag{2.6}$$

We say that $T_2^*(\cdot)$ is *Hadamard-continuous* at F if

$$|T_2^*(G) - T_2^*(F)| \to 0, \quad \text{with } \|G - F\| \to 0 \text{ on } G \in A. \tag{2.7}$$

Other regularity conditions will be introduced as and when needed.

Let now $X_1, \dots, X_n$ be n i.i.d.r.v.'s with a distribution function (d.f.) F. For simplicity, we assume that F is defined on the real line R $[= (-\infty, \infty)]$ and denote by

$$F_n(x) = n^{-1} \sum_{i=1}^{n} I(X_i \le x), \qquad x \in R, \tag{2.8}$$

where $I(A)$ stands for the indicator function of the set A. Then, corresponding to the parameter $\theta = T(F)$, we consider the estimator

$$T_n = T(F_n), \qquad n \ge 1. \tag{2.9}$$

To introduce the pseudovariables, we denote by

$$F_{n-1}^{(i)}(x) = (n-1)^{-1} \sum_{j=1\,(\ne i)}^{n} I(X_j \le x), \qquad x \in R,\ i = 1, \dots, n, \tag{2.10}$$

$$T_{n-1}^{(i)} = T\big(F_{n-1}^{(i)}\big) \quad \text{and} \quad T_{n,i} = nT_n - (n-1)T_{n-1}^{(i)}, \qquad i = 1, \dots, n. \tag{2.11}$$

Then the $T_{n,i}$ are the pseudovariables generated by jackknifing, and

$$T_n^* = n^{-1} \sum_{i=1}^{n} T_{n,i} \quad \text{and} \quad V_n^* = (n-1)^{-1} \sum_{i=1}^{n} (T_{n,i} - T_n^*)^2 \tag{2.12}$$

are the *classical jackknifed estimator* of $T(F)$ and the *jackknifed variance estimator*, respectively. To formulate the SOADR results for the classical jackknifing, first we denote

$$\overline{T}_{1n} = n^{-1} \sum_{i=1}^{n} T_1(F; X_i) = \int T_1(F; x)\, dF_n(x), \tag{2.13}$$

and note that [viz., Parr (1985)] whenever

$$0 < \sigma_1^2 = E_F\{T_1^2(F; X_1)\} < \infty, \tag{2.14}$$

the following holds:

$$n^{1/2}\big(T_n - T(F) - \overline{T}_{1n}\big) \to_p 0, \quad \text{as } n \to \infty, \tag{2.15}$$

$$n^{1/2}\big(T_n^* - T(F) - \overline{T}_{1n}\big) \to_p 0, \quad \text{as } n \to \infty, \tag{2.16}$$

$$n^{1/2}\overline{T}_{1n} \to_{\mathscr{D}} \mathscr{N}\big(0, \sigma_1^2\big). \tag{2.17}$$

Keeping these in mind, we define

$$\begin{aligned} R_n^* &= (n-1)(T_n - T_n^*), \\ R_n^{**} &= (n-1)\big(T_n^* - T(F) - \overline{T}_{1n}\big). \end{aligned} \tag{2.18}$$

R_n^* is essentially related to the *estimated bias* of T_n [see Parr (1985)], whereas R_n^{**} to the second-order representation for the classical jackknifing. We have the following.

THEOREM 2.1. *If $T(F)$ is second-order Hadamard-differentiable at F and $T_2^*(\cdot)$ is Hadamard-continuous at F, then*

$$R_n^* \to \tfrac{1}{2}T_2^*(F) \quad \textit{almost surely } (a.s.) \textit{ as } n \to \infty. \tag{2.19}$$

If $T(F)$ *is first-order Hadamard-differentiable at* F *and* $T_1^{**}(G) = \int T_1^2(G; x)\, dG(x)$ *is Hadamard-continuous at* F, *then* V_n^*, *defined by* (2.12), *converges a.s. to* σ_1^2 *as* $n \to \infty$.

REMARKS. (2.19) is comparable to Theorem 2 of Parr (1985). However, his strong second-order Fréchet differentiability condition seems to be more restrictive than the ones assumed here. Particularly, the Hadamard continuity seems to be very natural and easily verifiable than the extra regularity conditions in Parr (1985) needed to justify the "strong" part of the second-order Fréchet differentiability of $T(F)$. Also, (2.19) suggests that jackknifing in the classical case essentially amounts to a second-order bias adjustment without inducing any functional change in T_n. This also implies that T_n^* shares the same lack of robustness property with the initial estimator T_n when the later is not so robust. Finally, for the a.s. convergence of V_n^* to σ_1^2, it seems that we may as well replace the Parr (1985) "strong" first-order Fréchet differentiability by the ordinary first-order Hadamard differentiability and the Hadamard continuity of $T_1^{**}(\cdot)$, and this alternative setup is more easily verifiable.

To present the SOADR result on R_n^{**}, we assume that

$$E_F\{T_2^2(F; X_1, X_2)\} = \iint T_2^2(F; x, y)\, dF(x)\, dF(y) < \infty. \tag{2.20}$$

Then, from the basic results of Gregory (1977), we conclude that there exists a set (of finite or infinite collection of) *eigenvalues* $\{\lambda_k\}$ of $T_2(\cdot)$ corresponding to *orthonormal functions* $\{\tau_k(\cdot);\ k \geq 0\}$, such that

$$\int T_2(F; x, y)\tau_k(x)\, dF(x) = \lambda_k \tau_k(y) \quad \text{a.e. } (F), \forall\, k \geq 0, \tag{2.21}$$

$$\int \tau_k(x)\tau_q(x)\, dF(x) = \delta_{kq} \tag{2.22}$$

$$(= 1 \text{ or } 0 \text{ according as } k = q \text{ or not}),\ k, q \geq 0.$$

Note that the λ_k and $\tau_k(\cdot)$ may as well depend on F.

THEOREM 2.2. *Under* (2.2), (2.3), (2.7) *and* (2.20),

$$2R_n^{**} \to_{\mathscr{D}} \sum_{k \geq 0} \lambda_k (Z_k^2 - 1), \tag{2.23}$$

where the Z_k *are i.i.d.r.v.'s with the standard normal d.f.*

REMARK. The SOADR result for the classical jackknifing in (2.23) differs from the parallel result for M-estimators, considered by Jurečková (1985) and Jurečková and Sen (1987), among others. (2.23) is believed to be a novel and general SOADR result for classical jackknifing. It clearly reveals the role of the second-order compact derivative $T_2(F; \cdot)$ (and its eigenvalues $\{\lambda_k\}$) in the asymptotic distributional results of second order.

PROOF OF THEOREM 2.1. Note that, by (2.8) and (2.10),

$$\max_{1\le i\le n} \|F_{n-1}^{(i)} - F_n\| = \max_{1\le i\le n}\left\{\sup_x \left|F_{n-1}^{(i)}(x) - F_n(x)\right|\right\} = n^{-1}. \tag{2.24}$$

Furthermore, by (2.9) and (2.11),

$$T_{n,i} = T(F_n) + (n-1)\left\{T(F_n) - T\left(F_{n-1}^{(i)}\right)\right\}, \quad \text{for } i = 1,\dots,n. \tag{2.25}$$

Therefore, by (2.2)–(2.5) and (2.24)–(2.25), we have, for every i $(= 1,\dots,n)$,

$$\begin{aligned} T_{n,i} &= T_n + \int T_1(F_n; x)\, d\left[I(X_i \le x) - F_n(x)\right] \\ &\quad - \frac{1}{2(n-1)}\iint T_2(F_n; x, y)\, d\left[I(X_i \le x) - F_n(x)\right] \\ &\qquad \times d\left[I(X_i \le y) - F_n(y)\right] + o(n^{-1}) \\ &= T_n + T_1(F_n; X_i) - \frac{1}{2(n-1)} T_2(F_n; X_i, X_i) + o(n^{-1}), \end{aligned} \tag{2.26}$$

with probability 1. Thus, by (2.4)–(2.5), (2.12) and (2.26), we obtain

$$T_n^* = T_n - \frac{1}{2(n-1)}\int T_2(F_n; x, x)\, dF_n(x) + o(n^{-1}), \tag{2.27}$$

with probability 1, so that by (2.6), (2.18) and (2.27), we have

$$R_n^* = \tfrac{1}{2} T_2^*(F_n) + o(1). \tag{2.28}$$

Since $\|F_n - F\| \to 0$ a.s. as $n \to \infty$, invoking (2.7) on (2.28), we arrive at (2.19). Note that by using the first-order Hadamard differentiability of $T(F)$ along with (2.24), we readily obtain that with probability 1,

$$T_{n,i} - T_n^* = T_1(F_n; X_i) + o(1), \quad \text{for } i = 1,\dots,n. \tag{2.29}$$

Thus defining $T_1^{**}(\cdot)$ as in Theorem 2.1, we have

$$V_n^* = (n-1)^{-1} n T_1^{**}(F_n) + o(1) \quad \text{and} \quad \sigma_1^2 = T_1^{**}(F), \tag{2.30}$$

and hence the assumed Hadamard continuity of $T_1^{**}(\cdot)$ ensures the a.s. convergence of V_n^* to σ_1^2. □

PROOF OF THEOREM 2.2. By (2.2)–(2.5) and the fact that $\|F_n - F\| = O_p(n^{-1/2})$, we obtain

$$\begin{aligned} T_n &= T(F) + \int T_1(F; x)\, d\left[F_n(x) - F(x)\right] \\ &\quad + \tfrac{1}{2}\iint T_2(F; x, y)\, d\left[F_n(x) - F(x)\right] d\left[F_n(y) - F(y)\right] + o_p(n^{-1}) \\ &= T(F) + \overline{T}_{1n} + (2n)^{-1}\overline{T}_{2n} + (2n)^{-1}(n-1)U_n^{(2)} + o_p(n^{-1}), \end{aligned} \tag{2.31}$$

where $\overline{T}_{1n}$ is defined by (2.13), and

$$\overline{T}_{2n} = \int T_2(F; x, x)\, dF_n(x) = n^{-1} \sum_{i=1}^{n} T_2(F; X_i, X_i), \tag{2.32}$$

$$U_n^{(2)} = \binom{n}{2}^{-1} \sum_{\{1 \le i < j \le n\}} T_2(F; X_i, X_j). \tag{2.33}$$

Therefore, by (2.18), (2.27) and (2.31), we have

$$\begin{aligned} R_n^{**} &= (n-1)\{T_n^* - T_n + T_n - T(F) - \overline{T}_{1n}\} \\ &= -R_n^* + (n-1)(2n)^{-1}\overline{T}_{2n} + (1 - n^{-1})^2 (n/2) U_n^{(2)} + o_p(1). \end{aligned} \tag{2.34}$$

Now $\overline{T}_{2n}$, by (2.32), is an average over i.i.d.r.v.'s with finite first mean $T_2^*(F)$, and hence, by the Khintchine strong law of large numbers, $\overline{T}_{2n} \to T_2^*(F)$ a.s. as $n \to \infty$. Consequently, by (2.19), $(n-1)(2n)^{-1}\overline{T}_{2n} - R_n^* \to 0$ a.s. as $n \to \infty$. Furthermore, $U_n^{(2)}$ is a Hoeffding (1948) U-statistic with mean 0 [by (2.5)] and is *stationary of order 1* [by (2.5) and (2.20)]. Hence $|nU_n^{(2)}| = O_p(1)$. Thus, from (2.34), we have

$$R_n^{**} = (n/2)U_n^{(2)} + o_p(1). \tag{2.35}$$

Using the results of Gregory (1977) and Hall (1979), we have

$$P\{nU_n^{(2)} \le x\} \to P\left\{ \sum_{k \ge 0} \lambda_k (Z_k^2 - 1) \le x \right\}, \qquad x \in R, \tag{2.36}$$

so that (2.23) follows directly from (2.35) and (2.36). □

3. Functional jackknifing: Rationality. To motivate functional jackknifing, first, we denote the empirical d.f. of the pseudovariables by G_n, i.e., we let

$$G_n(x) = n^{-1} \sum_{i=1}^{n} I(T_{n,i} \le x), \qquad x \in R,\ n \ge 1. \tag{3.1}$$

Then, by (2.12) and (3.1),

$$\begin{aligned} T_n^* &= \int x\, dG_n(x), \\ V_n^* &= (n-1)^{-1} n \left\{ \int x^2\, dG_n(x) - \left(\int x\, dG_n(x) \right)^2 \right\}. \end{aligned} \tag{3.2}$$

For the particular case of $T_n = \overline{X}_n$, we have $G_n \equiv F_n$, so that $T_n^* = T_n = \overline{X}_n$ and $V_n^* = s_n^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$. However, for a general statistical functional $T(F)$, F_n and G_n are not generally equivalent, and, moreover, the $T_{n,i}$ are not independent. In fact, looking at (2.11), we may gather that because of the coefficients n and $n-1$ attached to T_n and $T_{n-1}^{(i)}$, the $T_{n,i}$ are more vulnerable to error contaminations (on the original X_i) and outliers. In such a case, though the appropriateness of the linear functional in (3.2) may be justified on the basis of the inherent reverse martingale structure of the resampling scheme in jackknifing [viz., Sen (1977)], on the ground of robustness and other considerations, other functionals of G_n appear to be more appealing. Indeed, Hinkley and Wang (1980) advocated the use of trimmed mean of the $T_{n,i}$, whereas Parr (1985), keeping in mind the equivalence of F_n and G_n for linear functionals, considered an L-functional of G_n (with a slight modification to achieve an $n^{-1/2}$ rate for the residual term, under more stringent conditions on the score function). Thus one may raise the issue in favor of a general functional

(3.3) $$T_n^0 = T^0(G_n), \quad \text{for a suitable } T^0(\cdot) \text{ defined on } D[0,1].$$

We term T_n^0 a *functional jackknifed estimator* (FJE) of $\theta = T(F)$.

Granted the existence of some G $(= G_F)$, such that $\|G_n - G\| \to_P 0$ (as $n \to \infty$), a minimal requirement for the rationality of T_n^0 as a suitable estimator of $T(F)$ is that

(3.4) $$T^0(G_F) = T(F), \quad \text{for all } F \text{ belonging to a class } \mathscr{F}.$$

In view of (3.2), we may as well set

(3.5) $$T(F) = \int x \, dG_F(x), \quad \text{for every } F \in \mathscr{F},$$

so that $T^0(\cdot)$ may be taken as some conventional functional related to the location model in the usual case. Thus functionals relating to R-, M- and L-estimators may be used. For any d.f. G, defined on R, we let $G(x; a) = G(x - a)$, for $a, x \in R$. Then a statistical functional $\tau(G)$ is said to be *translation-equivariant* (T.E.), if for every $a \in R$ and $G \in \mathscr{F}$,

(3.6) $$\tau(G(\cdot\,; a)) = a + \tau(G(\cdot\,; 0)).$$

It is clear that the functional T_n^* in (3.2) is T.E., and so are the other functionals considered by Hinkley and Wang (1980) and Parr (1985). We intend to retain this T.E. for our T_n^0 as well. For this, we define

(3.7) $$T_{n,i}^* = T_{n,i} - T_n^*, \quad \text{for } i = 1, \dots, n,$$

(3.8) $$G_n^*(x) = n^{-1} \sum_{i=1}^{n} I(T_{n,i}^* \le x) = G_n(x + T_n^*), \qquad x \in R.$$

Thus $\|G_n - G_F\| \to_P 0 \Rightarrow \|G_n^* - G_F^*\| \to_P 0$, where

(3.9) $$G_F^*(x) = G_F(x + T(F)), \qquad x \in R.$$

Note that by (2.4), $E_F T_1(F; X_1) = 0$, and, by (2.29), we may identify $G_F^*(x) = P\{T_1(F; X_1) \le x\}$, $x \in R$. Thus we may consider the following class of FJE:

(3.10) $$T^0(\cdot) \textit{ is translation-equivariant with } T^0(G_F^*) = 0.$$

For trimmed jackknifing, Hinkley and Wang (1980) assumed that $T_1(F; X_1)$ has a symmetric d.f. and this ensures (3.10), whereas for the classical jackknifing, (3.10) holds trivially as $E_F T_1(F; X_1) = 0$ [by (2.4)]. The symmetry of the d.f. of $T_1(F; X_1)$ (and its continuity) also suffice for an L-functional for $T^0(\cdot)$ [very close to what Parr (1985) suggested]; Parr's statistic may end up with estimating other forms of parameters if $T_1(F; X_1)$ does not have a symmetric and continuous d.f. In general, under this symmetry and continuity of the d.f. of $T_1(F; X_1)$, general (von Mises) functionals relating to R-, M- and L-estimators (of location) may be considered for $T^0(\cdot)$, and the specific choice of $T^0(\cdot)$ within this broad class may then be made on the ground of specific aspects of robustness, asymptotic minimaxity and other considerations. Our main contention is to present the general asymptotic theory of FJE [without restricting ourselves to specific subclasses of $T^0(\cdot)$], and, in the light of this theory, to make general comments on the scope as well as merits and demerits of FJE.

4. FJE: General asymptotics. Note that the consistency (in a weak or strong sense) of $T^0(G_n^*)$ (to 0) would ensure the same for T_n^0 (to θ). Similarly, the asymptotic behavior of $n^{1/2}T^0(G_n^*)$ dictates the asymptotic normality and other related results on T_n^0. Since, by construction, for every $n \geq 2$, $T_{n,1}^*, \ldots, T_{n,n}^*$ are exchangeable r.v.'s with $\sum_{i=1}^n T_{n,i}^* = 0$, we have $\int x\, dG_n^*(x) = 0$ a.e., so that for the classical jackknifing, there is no need to study the asymptotic behavior of $n^{1/2}T^0(G_n^*)$ (it is equal to 0 a.e.). For the trimmed jackknifing and general FJE, $n^{1/2}T^0(G_n^*)$ has a nondegenerate asymptotic distribution, and we intend to study the same. Toward this, we may find it convenient to incorporate the weak convergence of $n^{1/2}(G_n^* - G_F^*)$ (to an appropriate Gaussian function) along with plausible (first-order) Hadamard differentiability of $T^0(\cdot)$ in the formulation of the main results. However, as the $T_{n,1}^*$ are (generally) not independent, this weak convergence may not follow from the classical results (on empirical processes) and may need some extra regularity conditions.

We define a stochastic process $w_n = \{w_n(t),\ t \in [0,1]\}$ by letting

$$w_n(t) = n^{1/2}\{T_1(F_n; F_n^{-1}(t)) - T_1(F; F_n^{-1}(t))\}, \qquad t \in [0,1], \tag{4.1}$$

where $F_n^{-1}(t) = \inf\{x\colon F_n(x) \geq t\}$, $t \in [0,1]$ is the sample *quantile function.* Note that for the special case of $T_n = \bar{X}_n$, $w_n(t) = n^{1/2}(\bar{X}_n - E\bar{X}_n)$, for every $t \in [0,1]$ and is asymptotically normally distributed with 0 mean and variance $\sigma^2 = \mathrm{Var}(X_1)$. We assume that there exists a Gaussian function $\omega = \{\omega(t),\ t \in [0,1]\}$, such that

$$w_n \text{ converges in law to } \omega, \text{ as } n \to \infty. \tag{4.2}$$

We also assume that there exists an n_0, such that

$$T_2^2(F_n; X_1, X_1) \text{ is uniformly integrable, for } n \geq n_0. \tag{4.3}$$

In practice, both (4.2) and (4.3) can be verified by invoking standard techniques [when $T_1(\cdot)$ and $T_2(\cdot)$ are given], and these conditions appear to be less restrictive than Parr's (1985) strong second-order Fréchet differentiability of $T(F)$.

LEMMA 4.1. *Under* (4.3), *as* $n \to \infty$,

$$n^{-1/2}\Big\{\max_{1 \leq k \leq n} |T_2(F_n; X_k, X_k)|\Big\} \to_p 0. \tag{4.4}$$

PROOF. For every $\varepsilon > 0$,

$$
\begin{aligned}
&P\Big\{\max_{1\le k\le n}|T_2(F_n;X_k,X_k)|>\varepsilon\sqrt{n}\Big\}\\
&\quad\le \sum_{k=1}^{n} P\big\{|T_2(F_n;X_k,X_k)|>\varepsilon\sqrt{n}\big\}\\
&\quad\le (n\varepsilon^2)^{-1}\sum_{k=1}^{n} E\big\{T_2^2(F_n;X_k,X_k)I\big(T_2^2(F_n;X_k,X_k)>\varepsilon^2 n\big)\big\} \qquad (4.5)\\
&\quad= \varepsilon^{-2}E\big\{T_2^2(F_n;X_1,X_1)I\big(T_2^2(F_n;X_1,X_1)>\varepsilon^2 n\big)\big\}\\
&\quad\to 0, \quad \text{by (4.3).}
\end{aligned}
$$

□

Next, we note that by (2.26), (2.27), (3.7) and Lemma 4.1,

$$\max_{1\le k\le n}|T_{n,i}^{*}-T_1(F_n;X_i)| = o_p(n^{-1/2}), \quad \text{as } n\to\infty. \tag{4.6}$$

For later use, we denote the ordered values of the $T_{n,i}$ by $T_{n(i)}$, $i = 1,\dots,n$. Also, if $X_{n:1} \le \cdots \le X_{n:n}$ are the order statistics corresponding to $X_1,\dots,X_n$, then in (2.11) and (2.26), replacing the X_i by $X_{n:j}$, we denote the corresponding pseudovariables by $T_{n[j]}$, $j = 1,\dots,n$. Parr (1985) considered a version of the *pseudovariable quantile function* based on the $T_{n[j]}$, $j = 1,\dots,n$, whereas the natural version of the pseudovariable quantile function [i.e., $G_n^{-1}(t)$] is based on the $T_{n(i)}$, $i = 1,\dots,n$. Then, from (4.6), we arrive at the following.

LEMMA 4.2. *Whenever $T_1(F_n; x)$ is monotone in $x \in R$, under* (4.3), *the two pseudovariable quantile functions are $\sqrt{n}$-equivalent in probability i.e.,*

$$\max_{1\le k\le n}\big\{n^{1/2}|T_{n(k)}-T_{n[k]}|\big\} \to_p 0, \quad \text{as } n\to\infty. \tag{4.7}$$

Note that for a regular L-functional $T^0(\cdot)$, whenever $T_1(F; x)$ is monotone in x (otherwise, the L-functional may lose its rationality too), the results to follow would remain applicable to the jackknifed L-functional of Parr (1985), although his strong second-order Fréchet differentiability of $T(F)$ may not be that necessary.

We denote the true and empirical d.f.'s of the $T_1(F; X_i)$ by G_F^* and G_{n0}^*, respectively, so that

$$G_{n0}^*(x) = n^{-1}\sum_{i=1}^{n} I\big(T_1(F;X_i)\le x\big), \qquad x\in R,\ n\ge 1.$$

Also, we define $\{t_y;\ y \in R\}$ by letting $y = T_1(F; F^{-1}(t_y))$, $y \in R$, where, for an unessential simplification, we assume that $T_1(F; x)$ is monotone in x.

LEMMA 4.3. *If $T(F)$ is second-order Hadamard-differentiable at F, G_F^* is continuous a.e. and* (4.2) *and* (4.3) *hold, then*

$$\sup_y\Big\{n^{1/2}\big|G_n^*(y)-G_{n0}^*\big(y-n^{-1/2}w_n(t_y)\big)\big|\Big\} \to_p 0, \quad \text{as } n\to\infty. \tag{4.8}$$

PROOF. By virtue of (4.1), (4.2) and (4.6), writing $w_{ni} = n^{1/2}\{T_1(F_n; X_i) - T_1(F; X_i)\}$, $i = 1, \dots, n$, we have

$$\max_{1 \le i \le k} \left\{ n^{1/2} \left| T^*_{n,i} - T_1(F; X_i) - n^{-1/2} w_{ni} \right| \right\} \to_p 0, \quad \text{as } n \to \infty. \tag{4.9}$$

Furthermore, by virtue of (4.2), for every $\varepsilon > 0$ and $\eta > 0$, there exist positive constants K, n_0 and δ_0 $(0 < \delta_0 < 1)$ such that for every $n \ge n_0$,

$$P\left\{ \max_{1 \le i \le n} |w_{ni}| > K \right\} < \varepsilon, \tag{4.10}$$

$$P\left\{ \sup\left\{ |w_n(t) - w_n(s)| : 0 \le s \le t \le s + \delta \le 1 \right\} > \eta \right\} < \varepsilon, \quad \forall\, \delta \le \delta_0. \tag{4.11}$$

Also, $n^{1/2}\{G^*_{n0} - G^*_F\}$ converges weakly to a Gaussian function (reducible to a Brownian bridge), so that for every $n \ge n_0$,

$$P\left\{ \sup\left\{ n^{1/2} |G^*_{n0}(x) - G^*_{n0}(y) - G^*_F(x) + G^*_F(y)| : |x - y| < \delta' \right\} > \eta \right\} < \varepsilon, \tag{4.12}$$

where δ' (> 0) converges to 0 as $\delta \downarrow 0$.

For any given $y \in R$, consider a partition of R into $(-\infty, y - 2Kn^{-1/2})$, $[y - 2Kn^{-1/2}, y + 2Kn^{-1/2}]$ and $(y + 2Kn^{-1/2}, \infty)$. By (4.9) and (4.10), with probability $\ge 1 - \varepsilon$, we have (a) for all i such that $T_1(F; X_i) < y - 2Kn^{-1/2}$, $T^*_{n,i} < y - Kn^{-1/2} \le y + n^{-1/2} w_n(t_y)$, (b) for all i such that $T_1(F; X_i) > y + 2Kn^{-1/2}$, $T^*_{n,i} > y + n^{-1/2}K \ge y + n^{-1/2} w_n(t_y)$, and (c) [by (4.9) and (4.11)], for all i such that $y - 2n^{-1/2}K \ge T_1(F; X_i) \le y + 2n^{-1/2}K$, we have

$$T^*_{n,i} = T_1(F; X_i) + n^{-1/2} w_n(t_y) + o(n^{-1/2}). \tag{4.13}$$

Note that this picture holds uniformly in $y \in R$, and hence the rest of the proof of (4.8) follows by some standard arguments. □

LEMMA 4.4. *If* (i) $T(F)$ *satisfies the hypothesis of Lemma* 4.3 *and* (ii) $T^0(\cdot)$ *is T.E.* [*see* (3.10)] *and is first-order Hadamard-differentiable* (*at* G^*_F) *with the compact derivative* $T^0_1(G^*_F; y)$, *then*

$$n^{1/2}(T^0_n - T^*_n) = o_p(1) \Leftrightarrow n^{-1/2} \sum_{i=1}^{n} T^0_1(G^*_F; T^*_{n,i}) = o_p(1). \tag{4.14}$$

PROOF. Note that by (4.2) and Lemma 4.3, $n^{1/2}\|G^*_n - G^*_F\| = O_p(1)$, whereas $T^0(\cdot)$ is assumed to be first-order Hadamard-differentiable. Thus, by using (3.10), we have

$$\begin{aligned} n^{1/2}(T^0_n - T^*_n) &= n^{1/2} T^0(G^*_n) = n^{1/2}[T^0(G^*_n) - T^0(G^*_F)] \\ &= n^{1/2} \int T^0_1(G^*_F; y)\, d[G^*_n(y) - G^*_F(y)] + o_p(1) \\ &= n^{-1/2} \sum_{i=1}^{n} T^0_1(G^*_F; T^*_{n,i}) + o_p(1), \end{aligned} \tag{4.15}$$

as $\int T^0_1(G^*_F; y)\, dG^*_F(y) = 0$ [by (2.4)]. □

Note that for the classical jackknifing, the left-hand side of (4.15) is exactly equal to 0, so is the leading term on the right-hand side. In general, this may not be true, and to establish the asymptotic $\sqrt{n}$-equivalence (in probability) of the classical jackknife and the FJE, it may be easier to verify that the right-hand side relation in (4.14) holds. Note that for every n (≥ 2), $T_1^0(G_F^*; T_{n,i}^*)$, $1 \leq i \leq n$, are interchangeable r.v.'s and $\sum_{i=1}^n T_{n,i}^* = 0$ with probability 1. Hence we shall find it convenient to rewrite the leading term on the right-hand side of (4.15) as

$$n^{-1/2} \sum_{i=1}^{n} \{T_1^0(G_F^*; T_{n,i}^*) - c_n T_{n,i}^*\}, \tag{4.16}$$

where c_n is a suitable constant. For the classical jackknifing, $c_n = 1$ and (4.16) is equal to 0. For the FJE, keeping in mind the T.E. location functionals, we would have generally $T_1^0(G_F^*; T_{n,i}^*)$ a monotone function of $T_{n,i}^*$, and hence c_n may be chosen as the usual regression coefficient of the $T_1^0(G_F^*; T_{n,i}^*)$ on the $T_{n,i}^*$ (when the regression line is taken to have 0 intercept, because of the T.E.). Thus, for the trimmed jackknifing treated by Hinkley and Wang (1980), when we have (at each end) α-trimming, for some α: $0 < \alpha < \frac{1}{2}$, then we have $c_n = (1 - 2\alpha)^{-1}$ and (4.16) converges to 0 (in probability) when α is small. Note that the $T_{n,i}^*$ are interchangeable with an intraclass correlation of $-(n-1)^{-1}$. Also, whenever $T_1^0(G_F^*; x)$ is nondecreasing in $x \in R$, $T_1^0(G_F^*; T_{n,i}^*)$ and $T_{n,i}^*$ are positively associated. Furthermore, the $T_1^0(G_F^*; T_{n,i}^*)$ are also interchangeable r.v.'s, although their intraclass correlation may not be negative. It follows therefore that for any arbitrary c_n, the $T_1^0(G_F^*; T_{n,i}^*) - c_n T_{n,i}^*$ are interchangeable r.v.'s, and we may exploit this in the characterization of the stochastic equivalence of the classical and FJE.

LEMMA 4.5. *If there exists a sequence* $\{c_n\}$ *such that* (i) *the intraclass correlation of the* $T_1^0(G_F^*; T_{n,i}^*) - c_n T_{n,i}^*$ *is nonpositive, and* (ii) $E[T_1^0(G_F^*; T_{n,1}^*) - c_n T_{n,1}^*]^2$ *converges to* 0 *as* $n \to \infty$, *then* $n^{1/2}(T_n^0 - T_n^*)$ *converges to* 0 *in probability as* $n \to \infty$.

PROOF. Note that for nonpositive intraclass correlation, the second moment of the statistic in (4.16) is bounded from above by $E[T_1^0(G_F^*; T_{n,1}^*) - c_n T_{n,1}^*]^2$, and hence the desired result follows by using the Chebyshev inequality. □

If our basic goal is to choose a FJE such that some robustness is achieved without compromising on the asymptotic equivalence to the classical jackknifing, then Lemma 4.4 or 4.5 can be used with advantage to construct such $T^0(\cdot)$. In such a case, the asymptotic normality of the FJE also follows from that of the classical jackknifed estimator, and there is no need to consider variance estimators other than V_n^* in (2.12). However, in a general FJE, this picture may not hold and we need to carry out a more elaborate analysis.

Let us write $Y_i = T_1(F; X_i) + T_1^0(G_F^*; T_1(F; X_i))$, $i \geq 1$, and for every $n \geq 1$, we let

$$\begin{aligned} Y_{ni} &= n^{-1/2}[T_1(F; X_i) + T_1^0(G_F^*; T_{n,i}^*)], \\ v_{ni} &= T_1^0(G_F^*; T_{n,i}^*) - T_1^0(G_F^*; T_1(F; X_i)), \end{aligned} \tag{4.17}$$

for $i = 1, \dots, n$, so that $Y_{ni} = n^{-1/2}(Y_i + v_{ni})$, for $i = 1, \dots, n$; $n \geq 1$. Next, recall that

$$n^{1/2}(T_n^0 - T(F)) = n^{1/2}[T^0(G_n) - T(F)] = n^{1/2}[T_n^* - T(F) + T^0(G_n^*)]$$

$$(4.18) \qquad = n^{1/2}\overline{T}_{1n} + n^{-1/2}\sum_{i=1}^{n} T_1^0(G_F^*; T_{n,i}^*) + o_p(1)$$

$$= Y_{n1} + \cdots + Y_{nn} + o_p(1).$$

Note that for each n (≥ 1), $Y_{n1}, \dots, Y_{nn}$ are interchangeable (but not necessarily i.i.d.) r.v.'s; we denote by $\mathscr{F}_{n,k}$ the σ-field generated by $Y_{n,i}$; $i \leq k$, for $k = 0, \dots, n$ (where $\mathscr{F}_{0,n}$ is the trivial σ-field). We assume that as $n \to \infty$,

$$(4.19) \qquad \sum_{i=1}^{n} E[Y_{ni} | \mathscr{F}_{n,i-1}] \to_P 0,$$

$$(4.20) \qquad \sum_{i=1}^{n} \mathrm{Var}[Y_{ni} | \mathscr{F}_{n,i-1}] \to_P \gamma^2,$$

$$(4.21) \qquad \sum_{i=1}^{n} E[Y_{ni}^2 I(|Y_{ni}| > \varepsilon\gamma) | \mathscr{F}_{n,i-1}] \to_P 0, \quad \text{for every } \varepsilon > 0,$$

where γ^2 is a finite positive constant. Then, by the Dvoretzky (1972) central limit theorem (for a triangular scheme of possibly dependent r.v.'s), from (4.18)–(4.21), we arrive at the following.

THEOREM 4.6. *Suppose that the hypothesis of Lemma* 4.3 *holds and the* Y_{ni} *defined by* (4.17) *satisfy conditions* (4.19)–(4.21). *Then as* $n \to \infty$,

$$(4.22) \qquad n^{1/2}(T_n^0 - T(F)) \to_{\mathscr{D}} \mathscr{N}(0, \gamma^2).$$

The above theorem, formulated in a general fashion, rests on the verification of the three conditions in (4.19)–(4.21). In this context, note that the Y_i are i.i.d.r.v.'s, and note that $EY_i = 0$ whenever the compact derivatives $T_1(F; X_i)$ and $T_1^0(G_F^*; T_1(F; X_i))$ are integrable. Under similar square integrability conditions, we may assume that

$$\sigma_Y^2 = EY^2 = \int [T_1(F; x) + T_1^0(G_F^*; T_1(F; x))]^2 \, dF(x)$$

exists and is positive. Furthermore, by (4.6) and (4.9), we may write

$$(4.23) \qquad v_{ni} \doteq [T_1^0(G_F^*; T_1(F; X_i) + n^{-1/2}w_{ni} + \xi_{ni}) - T_1^0(G_F^*; T_1(F; X_i))], \quad i = 1, \dots, n,$$

where $w_n^* = \max\{|w_{ni}|: 1 \leq i \leq n\} = O_p(1)$ and $\xi_n^* = \max\{|\xi_{ni}|: 1 \leq i \leq n\} = o_p(1)$. Consequently, if $T_1^0(G_F^*; y)$ is equicontinuous (in y a.e.), then the v_{ni} are uniformly (in i) $O_p(n^{-1/2})$, so that (4.19)–(4.21) may easily be verified by reference to the Y_i, and for this the finiteness of γ^2 and the stochastic convergence of $n^{-1/2}\sum_{i=1}^{n} v_{ni}$ to 0 suffice. For the trimmed jackknifing of Hinkley and Wang (1980) as well as the trimmed L-functional jackknifing, this equicontinuity condition is easy to verify, and hence Theorem 4.6 applies directly under the usual Lindeberg–Feller condition on the Y_i. However, in general, for an unbounded functional, this equicontinuity condition may not hold, and, moreover, in general, $n^{-1/2}\sum_{i=1}^{n} v_{ni}$ may not converge to 0 (in probability), but may have a nondegenerate asymptotic distribution. In such a case, the verification of the three conditions in (4.19)–(4.21) may require more elaborate analysis. This additional complication can be avoided by an alternative approach wherein we impose some other (natural) regularity conditions on the compact derivative $T_1^0(\cdot)$. Toward this, we assume that $T_1^0(\cdot)$ admits the expansion

$$\begin{aligned} &T_1^0(G_F^*; T_1(F; X_i) + n^{-1/2}t) \\ &\quad = T_1^0(G_F^*; T_1(F; X_i)) + n^{-1/2}tT_{11}^0(G_F^*; T_1(F; X_i)) \\ &\qquad + o_p(n^{-1/2}), \quad \text{uniformly in } t\colon |t| \le T < \infty, \end{aligned} \tag{4.24}$$

where $T_{11}^0(\cdot)$ stands for the first derivative of $T_1^0(\cdot)$. Exploiting (4.9) and (4.10), we may then write the penultimate step in (4.22) as

$$\begin{aligned} T_n^0 - T(F) = \overline{T}_{1n} &+ n^{-1}\sum_{i=1}^{n} T_1^0(G_F^*; T_1(F; X_i)) \\ &+ n^{-3/2}\sum_{i=1}^{n} w_{ni}T_{11}^0(G_F^*; T_1(F; X_i)) \\ &+ o_p(n^{-1/2}). \end{aligned} \tag{4.25}$$

We note that by (2.4)–(2.5), the Hadamard derivative of $T_1(F; x)$ is given by $T_{1,1}(F; x, y) = T_2(F; x, y) - T_1(F; y)$, so that for every i $(= 1, \dots, n)$, we have

$$\begin{aligned} T_1(F_n; X_i) &= T_1(F; X_i) + \int T_{1,1}(F; X_i, y)\, d[F_n(y) - F(y)] + o(\|F_n - F\|) \\ &= T_1(F; X_i) + n^{-1}\sum_{j=1}^{n} T_2(F; X_i, X_j) \\ &\quad - n^{-1}\sum_{j=1}^{n} T_1(F; X_j) + o_p(n^{-1/2}). \end{aligned} \tag{4.26}$$

Using (4.26) for the w_{ni}, the right-hand side of (4.25) can be expressed as

$$\begin{aligned} &\overline{T}_{1n}\left\{1 - n^{-1}\sum_{j=1}^{n} T_{11}^0(G_F^*; T_1(F; X_j))\right\} + n^{-1}\sum_{i=1}^{n} T_1^0(G_F^*; T_1(F; X_i)) \\ &\quad + n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} T_2(F; X_i, X_j)T_{11}^0(G_F^*; T_1(F; X_i)) + o_p(n^{-1/2}). \end{aligned} \tag{4.27}$$

Next, we may note that by the Khintchine strong law of large numbers,

$$\begin{aligned} n^{-1}\sum_{i=1}^{n} T_{11}^0(G_F^*; T_1(F; X_i)) &\longrightarrow_{a.s.} \int T_{11}^0(G_F^*; T_1(F; x))\, dF(x) \\ &= v^*, \text{ say.} \end{aligned} \tag{4.28}$$

Note that by virtue of (3.10), for every real a, $T^0(G_F^*(\cdot; a)) = a + T^0(G_F^*) = a$, so that we have

$$\begin{aligned} a &= T^0(G_F^*(\cdot\,; a)) \\ &= T^0(G_F^*) + \int T_1^0(G_F^*; y)\, d[G_F^*(y-a) - G_F^*(y)] + o(a) \\ &= \int [T_1^0(G_F^*; y+a) - T_1^0(G_F^*; y)]\, dG_F^*(y) + o(a), \end{aligned}$$

and hence dividing both sides by a and allowing $a \to 0$, we immediately obtain $\int T_{11}^0(G_F^*; y)\, dG_F^*(y) = 1$. This implies that $v^* = 1$. Furthermore,

$$\begin{aligned} n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} & T_2(F; X_i, X_j) T_{11}^0(G_F^*; T_1(F; X_i)) \\ &= n^{-2} \sum_{i=1}^{n} T_2(F; X_i, X_i) T_{11}^0(G_F^*; T_1(F; X_i)) \\ &\quad + (n-1)n^{-1} \left\{ \binom{n}{2}^{-1} \sum_{\{1 \le i < j \le n\}} \phi(X_i, X_j) \right\} \\ &= n^{-1} U_{n(1)} + (n-1) n^{-1} U_{n(2)}, \quad \text{say}, \end{aligned} \tag{4.29}$$

where

$$U_{n(1)} \to v^{**} = E_F T_2(F; X_i, X_i) T_{11}^0(G_F^*; T_1(F; X_i)) \quad \text{almost surely}, \tag{4.30}$$

whenever the expectation exits, and $U_{n(2)}$ is a Hoeffding (1948) U-statistic of degree 2 corresponding to the kernel

$$\phi(X_i, X_j) = T_2(F; X_i, X_j)\left[T_{11}^0(G_F^*; T_1(F; X_i)) + T_{11}^0(G_F^*; T_1(F; X_j))\right]/2.$$

Thus if we assume that

$$E_F\left\{[\phi(X_i, X_j)]^2\right\} < \infty, \tag{4.31}$$

and denote by

$$\phi_1(x) = E_F \phi(x, X_1) = \tfrac{1}{2} \int T_2(F; x, y) T_{11}^0(G_F^*; T_1(F; y)\, dF(y), \qquad x \in R, \tag{4.32}$$

then, by the classical results of Hoeffding (1948), we have

$$n^{1/2}\left\{ U_{n(2)} - 2n^{-1} \sum_{i=1}^{n} \phi_1(X_i) \right\} \to_p 0, \quad \text{as } n \to \infty. \tag{4.33}$$

Consequently, if we define

$$\begin{aligned} \psi(x) &= T_1^0(G_F^*; T_1(F; x)) + 2\phi_1(x) \\ &= T_1^0(G_F^*; T_1(F; x)) + \int T_2(F; x, y) T_{11}^0(G_F^*; T_1(F; y))\, dF(y), \end{aligned} \tag{4.34}$$

then, by (4.27)–(4.34), we obtain

$$n^{1/2}\left\{ T_n^0 - T(F) - n^{-1} \sum_{i=1}^{n} \psi(X_i) \right\} \to_p 0, \quad \text{as } n \to \infty. \tag{4.35}$$

This leads us to the following.

THEOREM 4.7. *Suppose that* (4.24), (4.31) *and the hypothesis of Lemma* 4.3 *hold, and* v^{**} *defined by* (4.30) *is finite. Then for the FJE* T_n^0 *we have*

$$n^{1/2}\big(T_n^0 - T(F)\big) \to_{\mathscr{D}} \mathscr{N}\big(0, \sigma_0^2\big), \tag{4.36}$$

where

$$\sigma_0^2 = E_F\big[\{\psi(X_1)\}^2\big] \quad \text{and} \quad \psi(x) \text{ is defined by } (4.34). \tag{4.37}$$

It may be noted that for the classical jackknifing, $T_1^0(G_F^*; T_1(F; x)) = T_1(F; x)$, and hence $\psi(x) = T_1(F; x)$. Thus $\sigma_1^2 = \sigma_0^2$. Theorem 4.7 immediately leads us to the following.

COROLLARY 4.7.1. *Suppose that the conditions of Theorem* 4.7 *hold. Then the classical and FJE are square-root-n stochastically equivalent, whenever*

$$\begin{aligned} &T_1^0(G_F^*; T_1(F; x)) + \int T_2(F; x, y) T_{11}^0(G_F^*; T_1(F; y))\, dF(y) \\ &\quad = T_1(F; x) \quad a.e. \end{aligned} \tag{4.38}$$

This explains the role of the compact derivatives $T_1(\cdot)$, $T_2(\cdot)$, $T_1^0(\cdot)$ and the (partial) derivative $T_{11}^0(\cdot)$ [of $T_1^0(\cdot)$] in the maintenance of the asymptotic closeness [up to $O(n^{-1/2})$] of the classical and FJE, and in the endeavor of enhancing the robustness of FJE, we should keep (4.38) in mind so that we do not deviate too far. However, in general, for FJE (4.38) may not hold, and σ_0^2 defined by (4.37) is different from σ_1^2 defined by (2.14). Thus to make full use of FJE in drawing statistical inference on $T(F)$, we may need to estimate σ_0^2.

5. FJE: Estimation of asymptotic variance. In Theorem 2.1 we have established the a.s. convergence of the jackknifed variance V_n^* to σ_1^2 defined by (2.14). Also, in the last section, we have shown that for the FJE, the asymptotic normality holds with the asymptotic variance σ_0^2 defined by (4.37) and that σ_1^2 and σ_0^2 may not be equal. For a trimmed jackknifed estimator, Hinkley and Wang (1980) have suggested a suitable method of estimating σ_0^2, and Parr (1985) has also a suggestion in his case. For general FJE, we would like to consider a *two-step jackknifing* for the variance estimation.

Toward this proposal, we may note that $T_n^0 = T^0(G_n)$, where G_n is the empirical d.f. of the $T_{n,i}$, defined by (2.11). Let $T_{n-1}^{(i)}$ (and $T_{n-2}^{(ij)}$) be the statistic T_n computed from a sample of size $n-1$ (and $n-2$) obtained by deleting X_i (and X_i, X_j) from the given sample of size n, for $i \neq j = 1, \dots, n$. For each i ($= 1, \dots, n$), define

$$T_{n,i:j} = (n-1)T_{n-1}^{(i)} - (n-2)T_{n-2}^{(ij)}, \quad \text{for } j\ (\neq i) = 1, \dots, n. \tag{5.1}$$

If we denote the empirical d.f.'s for the samples of sizes $n-1$ and $n-2$ (resulting from the deletion of X_i and X_i, X_j from the complete sample of size n)

by $F_{n-1}^{(i)}$ and $F_{n-2}^{(ij)}$, then using the same expansions as in (2.24)–(2.27), we have

$$
\begin{aligned}
T_{n,i:j} &= T(F_{n-1}^{(i)}) - (n-2)[T(F_{n-2}^{(ij)}) - T(F_{n-1}^{(i)})] \\
&= T(F_n) + [T(F_{n-1}^{(i)}) - T(F_n)] - (n-2)[T(F_{n-2}^{(ij)}) - T(F_{n-1}^{(i)})] \\
&= T(F_n) + T_1(F_n; X_j) \\
&\quad - (2(n-2))^{-1}[2T_2(F_n; X_i, X_j) + T_2(F_n; X_j, X_j)] + r_{n,i:j},
\end{aligned} \tag{5.2}
$$

where

$$
\max_{1 \le i \ne j \le n} \{n|r_{n,i:j}|\} \to 0, \quad \text{a.s. as } n \to \infty. \tag{5.3}
$$

As such, by (2.26), (5.2) and (5.3), we obtain

$$
\max_{1 \le i \ne j \le n} \left|T_{n,i:j} - T_{n,j} + n^{-1}T_2(F_n; X_i, X_j)\right| = o(n^{-1}), \quad \text{a.s. as } n \to \infty. \tag{5.4}
$$

For each i $(= 1, \ldots, n)$, the empirical d.f. of the $T_{n,i:j}$ is denoted by $G_{n-1}^{(i)}$, whereas, as in Section 2, the empirical d.f. of the $T_{n,i}$ is denoted by G_n. Then, using (5.4) and proceeding as in the proof of Lemma 4.3, it follows that

$$
\max_{1 \le i \le n} \sup_x \left\{n^{1/2}|G_{n-1}^{(i)}(x) - G_n(x)|\right\} \to_p 0, \quad \text{as } n \to \infty. \tag{5.5}
$$

At the second stage of jackknifing, we identify that the FJE based on the $T_{n,i:j}$ $(j = 1, \ldots, n$ with $j \ne i)$ is nothing but $T^0(G_{n-1}^{(i)})$ for $i = 1, \ldots, n$. Thus the pseudovariables generated by these FJE are given by

$$
Q_{n,i} = nT^0(G_n) - (n-1)T^0(G_{n-1}^{(i)}), \quad \text{for } i = 1, \ldots, n. \tag{5.6}
$$

Using (5.4)–(5.6), we obtain

$$
\begin{aligned}
Q_{n,i} &= T^0(G_n) - (n-1)[T^0(G_{n-1}^{(i)}) - T^0(G_n)] \\
&= T^0(G_n) - (n-1)\int T_1^0(G_n; x)\, d[G_{n-1}^{(i)}(x) - G_n(x)] \\
&\quad + O(n\|G_{n-1}^{(i)} - G_n\|^2) \\
&= T^0(G_n) - (n-1)\int T_1^0(G_n; x)\, dG_{n-1}^{(i)}(x) + o_p(1) \\
&= T_n^0 - \sum_{j=1\,(j \ne i)}^{n} T_1^0(G_n; T_{n,i:j}) + o_p(1) \\
&= T_n^0 - \sum_{j=1\,(j \ne i)}^{n} \{T_1^0(G_n; T_{n,j}) - n^{-1}T_2(F_n; X_i, X_j)T_{11}^0(G_n; T_{n,j})\} + o_p(1) \qquad (5.7) \\
&= T_n^0 - n\int T_1^0(G_n; x)\, dG_n(x) + T_1^0(G_n; T_{n,i}) \\
&\quad + n^{-1}\sum_{j=1}^{n} T_2(F_n; X_i, X_j)T_{11}^0(G_n; T_{n,j}) \\
&\quad - n^{-1}T_2(F_n; X_i, X_i)T_{11}^0(G_n; T_{n,i}) + o_p(1) \\
&= T_n^0 + T_1^0(G_n; T_{n,i}) + n^{-1}\sum_{j=1}^{n} T_2(F_n; X_i, X_j)T_{11}^0(G_n; T_{n,j}) + o_p(1),
\end{aligned}
$$

where we assume that $T^0(\cdot)$ is second-order Hadamard-differentiable and the expansion in (4.24) holds. Thus, making use of (2.4)–(2.5), we obtain from (5.7) that

$$\bar{Q}_n = n^{-1} \sum_{i=1}^{n} Q_{n,i} = T_n^0 + 0 + 0 + o_p(1) = T_n^0 + o_p(1), \tag{5.8}$$

so that as $n \to \infty$,

$$\max_{1 \le i \le n} \left| \{Q_{n,i} - \bar{Q}_n\} - \left\{ T_1^0(G_n; T_{n,i}) + n^{-1} \sum_{j=1}^{n} T_2(F_n; X_i, X_j) T_{11}^0(G_n; T_{n,j}) \right\} \right| \to_p 0. \tag{5.9}$$

Note that by definition

$$n^{-1} \sum_{i=1}^{n} \{T_1^0(G_n; T_{n,i})\}^2 = \int \{T_1^0(G_n; x)\}^2 \, dG_n(x) = T_1^{0*}(G_n), \quad \text{say}. \tag{5.10}$$

It follows from our results in Sections 2 and 3 that $\|G_n - G_F\| \to_p 0$ as $n \to \infty$; G_F being the true d.f. of $T_1(F; X_1)$. Thus the Hadamard continuity [in the sense of (2.7)] ensures that as $n \to \infty$,

$$T_1^{0*}(G_n) \to_p T_1^{0*}(G_F) = T_1^{0*}(G_F^*), \tag{5.11}$$

where the last equality holds because of the translation-equivariance of $T^0(\cdot)$. A very similar treatment applies to the other two terms in the expansion of $n^{-1}\sum_{i=1}^{n}\{Q_{n,i} - \bar{Q}_n\}^2$, using only the leading terms in (5.9). Thus if we define

$$V_n^{**} = (n-1)^{-1} \sum_{i=1}^{n} (Q_{n,i} - \bar{Q}_n)^2, \tag{5.12}$$

we arrive at the following.

THEOREM 5.1. *Suppose that the hypothesis of Theorem* 4.7 *holds, and, in addition, the functionals in the expansion of* (5.12) *with the leading terms in* (5.9) *are all Hadamard-continuous. Then, defining* σ_0^2 *as in* (4.37), *we have*

$$V_n^{**} \to \sigma_0^2, \quad \textit{in probability as } n \to \infty. \tag{5.13}$$

Note that the construction of V_n^{**} is based on the FJE at the first step and the classical jackknifing at the second step. Thus V_n^{**} may be regarded as a two-step jackknifed variance estimator. This provides a natural jackknifed estimator of σ_0^2 and removes some of the arbitrariness in the alternative formulation of Hinkley and Wang (1980) and Parr (1985) for some particular cases.

6. Some general remarks. We may recall that for the classical jackknifing, the adjustment over the original estimator is $O(n^{-1})$ [see (2.19)]. On the other hand, for the FJE, such a strong result on the bias would require stronger regularity conditions. When seeking robustness through FJE, this refinement is of relatively minor importance [as the robustness adjustments are generally $O(n^{-1/2})$]. Thus a bias adjustment of $o(n^{-1/2})$ with a good robustness property of FJE may place it on a more attractive stand than the classical jackknifed estimator.

In Sections 3 and 4, we have mainly stressed the asymptotic normality of the classical and FJE. It is quite possible to extend the asymptotic normality results to parallel weak invariance principles for the partial sequence $\{n^{-1/2}k(T_k^0 - T(F)); k \le n\}$. A key to this invariance principle is provided by the well-known result on the empirical d.f. F_n

$$\max_{1 \le k \le n} \sup_x \left\{ n^{-1/2} k |F_k(x) - F(x)| \right\} = O_p(1). \tag{6.1}$$

As such, the results in Sections 3 and 4 may be extended in a routine manner.

The two-step jackknifing in Section 5 serves a very useful role in the estimation of σ_0^2. As has been explained earlier that, in general, σ_1^2 and σ_0^2 are not the same. Thus this difference reflects the relative increase in the asymptotic variance of the FJE (while attempting to induce more robustness). A comparison of V_n^* in (2.12) and V_n^{**} in (5.12) thus serves a useful role in the study of the robustness versus precision of the FJE. If T_n is asymptotically efficient, then $\sigma_0^2 \not< \sigma_1^2$. However, if T_n is not so, we may have even $\sigma_0^2 \le \sigma_1^2$, so that the FJE may induce robustness and enhance efficiency, too.

Acknowledgments. The author is grateful to the Associate Editor and the referees for their most useful comments on the manuscript.

REFERENCES

DVORETZKY, A. (1972). Central limit theorems for dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2** 513–555. Univ. California Press.

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

FERNHOLZ, L. T. (1983). *Von Mises Calculus for Statistical Functionals. Lecture Notes in Statist.* **19.** Springer, New York.

GREGORY, G. G. (1977). Large sample theory for *U*-statistics and tests of fit. *Ann. Statist.* **5** 110–123.

HALL, P. (1979). On the invariance principle for *U*-statistics. *Stochastic Process. Appl.* **9** 163–174.

HINKLEY, D. and WANG, H.-L. (1980). A trimmed jackknife. *J. Roy. Statist. Soc. Ser. B* **42** 347–356.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.

HUBER, P. J. (1981). *Robust Statistics.* Wiley, New York.

JUREČKOVÁ, J. (1985). Representation of *M*-estimators with the second-order asymptotic distribution. *Statist. Dec.* **3** 263–276.

JUREČKOVÁ, J. and SEN, P. K. (1987). A second order asymptotic distributional representation of *M*-estimators with discontinuous score functions. *Ann. Probab.* **15** 814–823.

MILLER, R. G., JR. (1974). An unbalanced jackknife. *Ann. Statist.* **2** 880–891.

PARR, W. C. (1983). A note on the jackknife, the bootstrap and the delta method estimators of bias and variance. *Biometrika* **70** 719–722.

PARR, W. C. (1985). Jackknifing differentiable statistical functionals. *J. Roy. Statist. Soc. Ser. B* **47** 56–66.

RAO, J. S. and SETHURAMAN, J. (1975). Weak convergence of empirical distribution functions of random variables subject to perturbations and scale factors. *Ann. Statist.* **3** 299–313.

SEN, P. K. (1977). Some invariance principles relating to jackknifing and their role in sequential analysis. *Ann. Statist.* **5** 316–329.

SEN, P. K. (1981). *Sequential Nonparametrics: Invariance Principles and Statistical Inference.* Wiley, New York.

DEPARTMENT OF BIOSTATISTICS 201H
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27514