The Annals of Statistics
1989, Vol. 17, No. 3, 1176–1197

A GENERAL THEORY FOR JACKKNIFE VARIANCE ESTIMATION[1]

BY JUN SHAO AND C. F. J. WU

Purdue University and University of Waterloo

The delete-1 jackknife is known to give inconsistent variance estimators for nonsmooth estimators such as the sample quantiles. This well-known deficiency can be rectified by using a more general jackknife with d, the number of observations deleted, depending on a smoothness measure of the point estimator. Our general theory explains why jackknife works or fails. It also shows that (i) for "sufficiently smooth" estimators, the jackknife variance estimators with bounded d are consistent and asymptotically unbiased and (ii) for "nonsmooth" estimators, d has to go to infinity at a rate explicitly determined by a smoothness measure to ensure consistency and asymptotic unbiasedness. Improved results are obtained for several classes of estimators. In particular, for the sample p-quantiles, the jackknife variance estimators with d satisfying $n^{1/2}/d \to 0$ and $n - d \to \infty$ are consistent and asymptotically unbiased.

1. Introduction. Variance estimators given by the delete-1 jackknife are known to be asymptotically consistent for sufficiently smooth estimators. If the estimator is not smooth, the jackknife may lead to an inconsistent variance estimator. The best known example of inconsistency is the sample quantile. See Miller (1974) for a review of both kinds of results. On the other hand, bootstrapping the sample quantile does lead to a consistent variance estimator under reasonable conditions on the underlying distribution [Efron (1982) and Ghosh, Parr, Singh and Babu (1984)]. This is a major triumph of the bootstrap over the jackknife. The main intent of this article is to remove this deficiency of the jackknife by proposing a more general version with d, the number of deleted observations in the jackknife, depending on a measure of smoothness of the estimator. The less smooth the estimator is, the larger d needs to be.

To aid our understanding of what causes inconsistency, let us examine more closely the example of the sample median, originally due to Moses. Let θ be the median of a distribution F, $X_1, \dots, X_n$ an i.i.d. sample from F and $X_{(1)} \le \cdots \le X_{(n)}$ their order statistics. Assume $n = 2m$. The estimator $\hat{\theta}$ is the sample median $(X_{(m)} + X_{(m+1)})/2$. Let $\hat{\theta}_{-i}$ be the median estimate after deleting X_i from the sample, i.e., $\hat{\theta}_{-i} = X_{(m+1)}$ for $i \le m$ and $X_{(m)}$ for $i \ge m+1$. In this case $\hat{\theta}$ is identical to the average of $\hat{\theta}_{-i}$. The jackknife variance estimator

$$v_{J(1)} = \frac{n-1}{n} \sum_1^n \left(\hat{\theta}_{-i} - \hat{\theta}\right)^2 \tag{1.1}$$

Received November 1986; revised July 1988.

[1]Research completed at the University of Wisconsin, Madison and supported by NSF-AFOSR Grants DMS-85-02303 and ISSA-86-0068.

AMS 1980 *subject classifications*. Primary 62G05; secondary 62E20, 62G99.

Key words and phrases. Asymptotic unbiasedness, balanced subsampling, consistency, Fréchet differentiability, grouped jackknife, L-estimator, M-estimator, sample quantile, smoothness of an estimator, U-statistic, von Mises expansion.

turns out to be

$$(n-1)\left(X_{(m+1)} - X_{(m)}\right)^2/4$$

and

$$nv_{J(1)} \to \sigma^2\left(\chi_2^2/2\right)^2 \quad \text{in distribution,}$$

where σ^2 is the asymptotic variance of $n^{1/2}(\hat{\theta} - \theta)$ and χ_2^2 is a chi-square random variable with 2 degrees of freedom [see Efron (1982), Chapter 3]. Hence $v_{J(1)}$ is inconsistent.

Intuitively $v_{J(1)}$ would work if $\hat{\theta}_{-i} - \hat{\theta}$ could mimic the sampling behavior of $\hat{\theta} - \theta$. However, $\hat{\theta}_{-i} - \hat{\theta} = O_p(n^{-1})$ is an order lower than $\hat{\theta} - \theta = O_p(n^{-1/2})$. For sufficiently smooth $\hat{\theta}$, this *mismatch* of orders does not cause problems since the scale factor $(n-1)$ in $v_{J(1)}$ will correct it. For less smooth $\hat{\theta}$ such as the sample quantiles, this simple correction by rescaling does not work. A remedy is to delete more observations, say d, in the repeated evaluations of $\hat{\theta}$. This method is called the delete-d jackknife. Formally we define the delete-d jackknife variance estimators as follows.

Suppose that $X_1, \ldots, X_n$ are i.i.d. with distribution $F \in \Xi$ and $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ is an estimator of a parameter θ. Here Ξ denotes the space of distributions. Without loss of generality, $\hat{\theta}$ is assumed to be symmetric. For a fixed n, let $d = d_n$ be an integer less than n and $r = n - d$. Define $\mathbf{S}_{n,r}$ to be the collection of subsets of $\{1, \ldots, n\}$ which have size r. For any $s = \{i_1, \ldots, i_r\} \in \mathbf{S}_{n,r}$, let $\hat{\theta}_s = \hat{\theta}(X_{i_1}, \ldots, X_{i_r})$. The delete-$d$ jackknife estimator of $\operatorname{Var}\hat{\theta}$ is

$$v_{J(d)} = \frac{r}{dN}\sum_s \left(\hat{\theta}_s - \hat{\theta}\right)^2, \tag{1.2}$$

where $N = \binom{n}{d}$ and Σ_s is the summation over all the subsets in $\mathbf{S}_{n,r}$. The delete-d jackknife was considered for other purposes, e.g., Bhargava (1983) and Wu (1986).

Another version of the delete-d jackknife variance estimator is

$$\tilde{v}_{J(d)} = \frac{r}{dN}\sum_s \left(\hat{\theta}_s - \frac{1}{N}\sum_s \hat{\theta}_s\right)^2 = v_{J(d)} - \frac{r}{d}\left(\frac{1}{N}\sum_s \hat{\theta}_s - \hat{\theta}\right)^2. \tag{1.3}$$

For certain estimators such as the U-statistics and the sample median, $\hat{\theta} = N^{-1}\Sigma_s\hat{\theta}_s$ and $v_{J(d)} = \tilde{v}_{J(d)}$. In general, $v_{J(d)}$ and $\tilde{v}_{J(d)}$ are unequal but are asymptotically equivalent iff

$$\frac{r}{d}\left(\frac{1}{N}\sum_s \hat{\theta}_s - \hat{\theta}\right)^2 = o_p(n^{-1}). \tag{1.4}$$

Sufficient conditions for (1.4) are given in Proposition 1 and Corollary 3.

When $\hat{\theta}$ is linear, i.e.,

$$\hat{\theta} = \theta + \frac{1}{n}\sum_1^n \phi_F(X_i), \tag{1.5}$$

for some function ϕ_F, it is easy to show that these estimators are identical to the usual variance estimator

$$(1.6) \qquad v_{J(d)} = \tilde{v}_{J(d)} = v_{J(1)} = \frac{1}{n(n-1)} \sum_1^n \left[\phi_F(X_i) - \frac{1}{n} \sum_1^n \phi_F(X_i) \right]^2 \quad \text{for any } d.$$

Using the rationale of "order matching," by choosing $d = \lambda n$, $0 < \lambda < 1$, in $v_{J(d)}$ or $\tilde{v}_{J(d)}$, the "resampling error" $\hat{\theta}_s - \hat{\theta}$ has $O_p(n^{-1/2})$ which matches the order of the "sampling error" $\hat{\theta} - \theta$. Moreover, the coefficient r/d in $v_{J(d)}$ is $O(1)$. Unlike other choices of d which result in $r/d \to 0$ or ∞, no drastic rescaling is required. It turns out that for this choice of d, $v_{J(d)}$ is consistent for a wide class of $\hat{\theta}$ (Corollary 1).

The special case of deleting half samples deserves particular attention. Here $d = n/2$, $r/d = 1$ and therefore no rescaling in $v_{J(d)}$ is required. In this regard $v_{J(d)}$ with $d = n/2$ behaves like the bootstrap variance estimator with B bootstrap samples [Efron (1982)]

$$v_b = \frac{1}{B} \sum_1^B \left(\hat{\theta}^{*b} - \hat{\theta}\right)^2, \qquad \hat{\theta}^{*b} = \hat{\theta}\left(X_1^{*b}, \dots, X_n^{*b}\right),$$

where the bth bootstrap sample $X_1^{*b}, \dots, X_n^{*b}$ is drawn at random with replacement from $X_1, \dots, X_n$. Both $\hat{\theta}^{*b} - \hat{\theta}$ and $\hat{\theta}_s - \hat{\theta}$ have the same order $O_p(n^{-1/2})$ as that of $\hat{\theta} - \theta$ and neither variance estimator requires rescaling. Half-sample estimates are commonly used in survey sampling [Rao and Wu (1985)].

Our approach was originally motivated by the above concept of "order matching," which, however, is not precise enough to treat cases in which order matching does not hold. For example, in the case of the sample quantile, it is found (Section 5, Example 1) that any d satisfying $n^{1/2}/d \to 0$, but not necessarily of the form $d = \lambda n$, will make $v_{J(d)}$ consistent. If the order of d is strictly between n and $n^{1/2}$, $\hat{\theta}_s - \hat{\theta}$ is of a lower order than $\hat{\theta} - \theta$ and yet consistency holds.

A more general concept than order matching for studying the asymptotic properties of $v_{J(d)}$ is given by the notion of "smoothness" of $\hat{\theta}$. Assume that $\hat{\theta}$ admits the expansion

$$(1.7) \qquad \hat{\theta} = \theta + \frac{1}{n} \sum_1^n \phi_F(X_i) + R_n,$$

where the function ϕ_F has mean zero and positive variance σ^2 and R_n is the remainder term. An important example of (1.7) is the von Mises expansion [see Serfling (1980)]. Since $n^{-1}\Sigma\phi_F(X_i)$ in (1.7) is $O_p(n^{-1/2})$, it is reasonable to assume that $R_n = o_p(n^{-1/2})$. We use the order of nER_n^2 as a *smoothness measure* of $\hat{\theta}$. Another smoothness measure is given in (3.2). Typically, $ER_n^2 = o(n^{-1})$. Necessary and sufficient conditions for this to hold are given in Lemma 1.

In Theorem 1 we establish a general condition on d in terms of ER_n^2 and n for $v_{J(d)}$ to be consistent and asymptotically unbiased. Necessary conditions for the consistency and asymptotical unbiasedness of $v_{J(d)}$ are established in Theorem 3. For sufficiently smooth $\hat{\theta}$, $v_{J(1)}$, and more generally $v_{J(d)}$ with bounded d, are consistent (Corollary 2 and Theorems 4 and 5). Theorem 2 provides further insights on why the choice of d in Theorem 1 makes $v_{J(d)}$ consistent. Noting that the expansion (1.7) decomposes $\hat{\theta} - \theta$ into the linear part $n^{-1}\Sigma\phi_F(X_i)$ and the nonlinear part R_n, $v_{J(d)}$ can be similarly decomposed into a linear component (i.e., $v_{J(d)}$ with $\hat{\theta}$ replaced by its linear part), a nonlinear component and their cross product. If the linear component dominates the nonlinear component, the asymptotic behavior of $v_{J(d)}$ is determined by its linear component and standard results apply. Precise conditions for this domination are given in Theorem 2 and the discussion following it. For sufficiently smooth $\hat{\theta}$, the linear component dominates even for $d = 1$. This explains the consistency of $v_{J(1)}$. For less smooth $\hat{\theta}$, the linear component may not dominate if d is bounded. The situation can be rectified by choosing a larger d to increase the order of the linear component so that it becomes dominant.

To alleviate the computational burden of $v_{J(d)}$ for $d > 1$, a more economic resampling plan, called balanced subsampling, is introduced in Section 5. All the results for $v_{J(d)}$ and $\tilde{v}_{J(d)}$ hold for this plan. Several examples, including the sample quantiles, L- and M-estimators, U-statistics and Fréchet differentiable functionals, are given in Section 6 to illustrate the theory. For the sample quantiles, $v_{J(d)}$ is shown to be consistent if d satisfies $n^{1/2}/d \to 0$ and F'' exists in a neighborhood of θ. Some concluding remarks are given in Section 7.

2. A technical lemma. Several results to be given later are obtained under the following condition on the remainder term R_n:

$$ER_n^2 = o(n^{-1}). \tag{2.1}$$

Lemma 1 states conditions equivalent to (2.1) that may be easier to verify in some situations. These conditions are combinations of the following:

$$R_n = o_p(n^{-1/2}), \tag{2.2}$$

$$E\left[\phi_F(X_1)R_n\right] = o(n^{-1}), \tag{2.3}$$

$$E\hat{\theta} = \theta + o(n^{-1/2}), \tag{2.4}$$

$$\operatorname{Var}\hat{\theta} = \frac{\sigma^2}{n} + o(n^{-1}), \tag{2.5}$$

$$\left\{n(\hat{\theta} - \theta)^2\right\} \text{ is uniformly integrable.} \tag{2.6}$$

Condition (2.2) is quite weak, since $n^{-1}\Sigma\phi_F(X_i)$ in (1.7) is $O_p(n^{-1/2})$ and (2.2) means that the remainder term R_n is of a lower order. When $\hat{\theta}$ is a statistical functional $T(F_n)$, where F_n is the empirical distribution of $X_1, \dots, X_n$, a sufficient condition for (2.2) is that T is quasi-Fréchet differentiable with respect to a norm $\|\ \|$ for which $\|F_n - F\| = O_p(n^{-1/2})$ [Serfling (1980), page 221]. Examples will be given later. Condition (2.4) is also quite weak. It means that the bias of $\hat{\theta}$

can be of any rate lower than $n^{-1/2}$. Typically the bias of $\hat{\theta}$ is of the order n^{-1} [Lehmann (1983), Section 2.5].

Under (2.2), σ^2/n is the variance of the asymptotic distribution of $\hat{\theta}$. Condition (2.5) ensures that σ^2/n is a valid asymptotic approximation to $\operatorname{Var}\hat{\theta}$. Usually $\operatorname{Var}\hat{\theta} \geq \sigma^2/n$ in view of Lemma 5.1.2 of Lehmann (1983) and $\operatorname{Var}\hat{\theta} = \sigma^2/n + o(n^{-1})$ for sufficiently regular estimators with finite variances. A limitation of our approach is that for some estimators (2.5) fails or is technically difficult to verify. This will be further discussed in Section 7(ii).

Sufficient conditions for (2.6) are given in Serfling (1980), pages 13–15. One such condition is

$$\sup_n E\left|n^{1/2}(\hat{\theta} - \theta)\right|^{2+\delta} < \infty \quad \text{for some } \delta > 0.$$

LEMMA 1. *The following four conditions are equivalent*:

(a) (2.1) *holds*.
(b) (2.2) *and* (2.5) *hold*.
(c) (2.3)–(2.5) *hold*.
(d) (2.2) *and* (2.6) *hold*.

PROOF. It is obvious that (2.1) implies (2.2), (2.4) and (2.5). From

$$n \operatorname{Var}\hat{\theta} = \sigma^2 + n \operatorname{Var} R_n + 2nE\left[\phi_F(X_1)R_n\right], \tag{2.7}$$

(2.1) implies (2.3). Therefore (a) implies (b) and (c). To show that (c) implies (a), note that (2.4) is equivalent to $ER_n = o(n^{-1/2})$ and (2.3) and (2.5) imply $\operatorname{Var} R_n = o(n^{-1})$ via (2.7). Hence (a) is equivalent to (c). To show that (b) implies (d), note that

$$nR_n^2 \leq 2\left[n(\hat{\theta} - \theta)^2 + \frac{1}{n}\left(\sum_1^n \phi_F(X_i)\right)^2\right] \tag{2.8}$$

and

$$n^{-1}E\left[\sum_1^n \phi_F(X_i)\right]^2 = \sigma^2. \tag{2.9}$$

Conditions (2.2) and (2.5) imply that $nE(\hat{\theta} - \theta)^2$ is bounded [see Lemma 2 of Ghosh, Parr, Singh and Babu (1984)], which together with (2.8) and (2.9) implies the uniform integrability of $n^{1/2}R_n$. Hence from (2.2), (2.4) follows. This and (2.5) imply

$$nE(\hat{\theta} - \theta)^2 \to \sigma^2. \tag{2.10}$$

From the central limit theorem,

$$n^{-1/2}\sum_1^n \phi_F(X_i) \to N(0, \sigma^2) \quad \text{in distribution,} \tag{2.11}$$

which, along with (2.2), implies that

$$n^{1/2}(\hat{\theta} - \theta) \to N(0, \sigma^2) \quad \text{in distribution.} \tag{2.12}$$

Now (2.10) and (2.12) imply that $\{n(\hat{\theta} - \theta)^2\}$ is uniformly integrable [Serfling (1980), page 15]. It remains to be shown that (d) implies (a). From (2.9) and (2.11), $n^{-1}[\sum_1^n \phi_F(X_i)]^2$ is uniformly integrable. Hence nR_n^2 is uniformly integrable by (2.6) and (2.8). Thus (a) follows from (2.2). This completes the proof. □

3. Consistency and asymptotic unbiasedness of jackknife variance estimators.

Several definitions are needed before stating the main results.

DEFINITION 1. For any estimator v of $\operatorname{Var}\hat{\theta} < \infty$:

(i) v is *consistent* for $\operatorname{Var}\hat{\theta}$ iff $(v - \operatorname{Var}\hat{\theta})/\operatorname{Var}\hat{\theta} = o_p(1)$.
(ii) v is *asymptotically unbiased* for $\operatorname{Var}\hat{\theta}$ iff $(Ev - \operatorname{Var}\hat{\theta})/\operatorname{Var}\hat{\theta} = o(1)$.

Definition 1 requires that $\operatorname{Var}\hat{\theta}$ be finite. A more general definition of consistency, even when $\operatorname{Var}\hat{\theta}$ does not exist, is to replace $\operatorname{Var}\hat{\theta}$ in Definition 1 by σ^2/n, the limiting variance of $\hat{\theta}$. That is, v is consistent iff $nv - \sigma^2 = o_p(1)$ and v is asymptotically unbiased iff $nEv - \sigma^2 = o(1)$. If condition (2.5) holds, the two definitions are equivalent.

Let $R_{n,s}$ be the remainder term in the expansion (1.7) for $\hat{\theta}_s$,

$$\hat{\theta}_s = \theta + \frac{1}{r}\sum_{i\in s}\phi_F(X_i) + R_{n,s}$$

and

$$U_s = R_{n,s} - R_n. \tag{3.1}$$

Since $\{X_i\}$ are exchangeable, $ER_{n,s}^2 = ER_r^2$ and $EU_s^2 = E(R_r - R_n)^2$ for any s, where $R_r = R_{n,s}$ with $s = \{1, \dots, r\}$. Write

$$\delta_r = rER_{n,s}^2 \quad \text{and} \quad \tau_r = rEU_s^2. \tag{3.2}$$

Note that δ_r is r times the mean squared error of the linear approximation to $\hat{\theta}_s$ with sample size r and τ_r is r times the expected mean squared deviation of $R_{n,s}$ from R_n. Both can be taken as measures of smoothness of $\hat{\theta}$. In this section, the asymptotic properties of $v_{J(d)}$ are studied by relating d to the smoothness measures δ_r and τ_r. All the results hold also for $\tilde{v}_{J(d)}$ with the smoothness measure τ_r replaced by

$$\tilde{\tau}_r = rEV_s^2, \qquad V_s = R_{n,s} - N^{-1}\sum_s R_{n,s}.$$

The proofs are similar and omitted.

The following theorem gives sufficient conditions for the consistency and asymptotic unbiasedness of $v_{J(d)}$.

Theorem 1. *Let $d = d_n$ be the number of observations deleted, $r = n - d$ and δ_r and τ_r be defined in* (3.2). *Assume either*

$$\lim_{n\to\infty} \frac{n\delta_r}{d} = 0 \tag{3.3}$$

or

$$\lim_{n\to\infty} \frac{n\tau_r}{d} = 0. \tag{3.4}$$

Then $v_{J(d)}$ is consistent and asymptotically unbiased.

Remarks. (i) Condition (3.3) implies $\lim_{n\to\infty} \delta_r = 0$, which is equivalent to (2.1) since $r \to \infty$ as $n \to \infty$.

(ii) Condition (3.4) is also necessary for the consistency and asymptotic unbiasedness of $v_{J(d)}$ (see Theorem 3). Condition (3.3) is stronger than (3.4) but is easier to check. Also (3.3) allows the choice of d to be made directly from the second moment of R_n without involving the subsets.

(iii) For the validity of the theorem, only the weak law of large numbers is needed in (3.6). The strong version of (3.6) will be used in the proof of Theorem 4.

Proof of Theorem 1. Let

$$L_s = \frac{1}{r} \sum_{i\in s} \phi_F(X_i) - \frac{1}{n}\sum_1^n \phi_F(X_i). \tag{3.5}$$

From (1.6) and the strong law of large numbers,

$$E\left(\frac{nr}{dN}\sum_s L_s^2\right) = \sigma^2 \quad \text{and} \quad \frac{nr}{dN}\sum_s L_s^2 \to \sigma^2 \text{ a.s.} \tag{3.6}$$

From (3.1) and (3.5),

$$nv_{J(d)} = \frac{nr}{dN}\sum_s (L_s + U_s)^2.$$

By the Cauchy–Schwarz inequality, the results follow from

$$E\left(\frac{nr}{dN}\sum_s U_s^2\right) = o(1),$$

which is guaranteed by (3.3) or (3.4). □

From (1.6), $v_{J(d)}$ for linear $\hat{\theta}$ is unbiased and consistent for any d. For nonlinear $\hat{\theta}$, Theorem 1 proves the asymptotic unbiasedness and consistency of $v_{J(d)}$ with a restriction on d, which depends on the smoothness of $\hat{\theta}$. Conditions (3.3) and (3.4) in Theorem 1 relate d to δ_r and τ_r, respectively. Thus the choice of d can be made via these relations, especially (3.3). For asymptotic unbiasedness and consistency d/n, the percentage of observations deleted in the jackknife, should be of a larger order than τ_r or δ_r. The smoother $\hat{\theta}$ is (i.e., smaller τ_r or δ_r), the smaller number of deleted observations is required.

Two important special cases are d bounded and d satisfying

$$\frac{d}{n} \geq \varepsilon_0 \quad \text{for some } \varepsilon_0 > 0 \quad \text{and} \quad r = n - d \to \infty. \tag{3.7}$$

The case of bounded d will be discussed in the next section. If d is chosen according to (3.7), (3.3) holds under the mild condition (2.1), i.e., $\delta_r \to 0$. This is stated as Corollary 1.

COROLLARY 1. *Under* (2.1), $v_{J(d)}$ *for* d *satisfying* (3.7) *is consistent and asymptotically unbiased.*

The interplay between the smoothness of $\hat{\theta}$ and the number of observations deleted in $v_{J(d)}$ is further explored in the following theorem.

THEOREM 2. *Assume that* (2.2) *holds. Let* $\{s_n\}$ *be any sequence of subsets such that* $s_n \in \mathbf{S}_{n,r}$. *For each* n, *let* L_{s_n} *and* U_{s_n} *be defined in* (3.5) *and* (3.1) *and*

$$e_{s_n} = \hat{\theta}_{s_n} - \hat{\theta} = L_{s_n} + U_{s_n}.$$

(i) *If* $d/n \to 0$ (*with* d *bounded or unbounded*), *then*

$$L_{s_n} \equiv O_p(d^{1/2} n^{-1}) \quad \text{and} \quad U_{s_n} = o_p(n^{-1/2}),$$

where $A_n \equiv O_p(a_n)$ *means that* A_n *has exactly the order* a_n.

(ii) *If there is an* $\varepsilon_0 > 0$ *such that*

$$\varepsilon_0 \leq \frac{d}{n} \leq 1 - \varepsilon_0, \tag{3.8}$$

then

$$L_{s_n} \equiv O_p(n^{-1/2}) \quad \text{and} \quad U_{s_n} = o_p(n^{-1/2}).$$

(iii) *If* $d/n \to 1$ *but* $r \to \infty$, *then*

$$L_{s_n} \equiv O_p(r^{-1/2}) \quad \text{and} \quad U_{s_n} = o_p(r^{-1/2}).$$

PROOF. We drop the subscript n in s_n in the proof. For any of the three cases, $U_s = o_p(r^{-1/2})$ by (2.2). Hence the assertion for U_s in (i)–(iii) is proved. Note that

$$L_s = \frac{d}{n}(L_{1,s} - L_{2,s}),$$

where

$$L_{1,s} = \frac{1}{r} \sum_{i \in s} \phi_F(X_i) \quad \text{and} \quad L_{2,s} = \frac{1}{d} \sum_{i \in \bar{s}} \phi_F(X_i)$$

are independent and $\bar{s}$ is the complement of s.

CASE (i). For bounded d, $n^{-1}\,dL_{1,s} \equiv O_p(n^{-3/2})$ and $n^{-1}\,dL_{2,s} \equiv O_p(n^{-1})$. Hence $L_s \equiv O_p(n^{-1})$. For $d \to \infty$ and $d/n \to 0$ (and therefore $r/n \to 1$), $n^{-1}\,dL_{1,s} \equiv O_p(dn^{-3/2})$ and $n^{-1}\,dL_{2,s} \equiv O_p(d^{1/2}n^{-1})$. Hence $L_s \equiv O_p(d^{1/2}n^{-1})$.

CASE (ii). From (3.8) and the independence of $L_{1,s}$ and $L_{2,s}$,

$$L_s \equiv \frac{d}{nr^{1/2}}O_p(1) + \frac{d^{1/2}}{n}O_p(1) \equiv O_p(n^{-1/2}).$$

CASE (iii). Since $d/n \to 1$ and $r \to \infty$, $r/n \to 0$, $L_{1,s} \equiv O_p(r^{-1/2})$ and $L_{2,s} \equiv O_p(d^{-1/2})$. Hence $L_s \equiv O_p(r^{-1/2})$. □

The quantity e_{s_n} can be interpreted as a "resampling error" and L_{s_n} (3.5) as its linear component. If $\hat{\theta}$ is linear in the sense of (1.5), then $e_{s_n} = L_{s_n}$. The previous theorem shows that if $d/n \geq \varepsilon_0 > 0$ and $r \to \infty$, the linear component of e_{s_n} is the dominating term and the jackknife variance estimator $v_{J(d)}$ behaves asymptotically as in the linear case. This explains why in Corollary 1, $v_{J(d)}$ is asymptotically unbiased and consistent without requiring any further smoothness condition on the point estimator $\hat{\theta}$ other than (2.1).

On the other hand, if $d/n \to 0$, the linear component of e_{s_n} is not necessarily dominant since from Theorem 2(i),

$$L_{s_n} \equiv O_p(d^{1/2}n^{-1}) \equiv \left(\frac{d}{n}\right)^{1/2} O_p(n^{-1/2}),$$

which may converge to zero faster than U_{s_n}. This explains why $v_{J(d)}$ with small d may not work well for nonsmooth $\hat{\theta}$. L_{s_n} becomes the dominating term iff

$$U_{s_n} = o_p\left(d^{1/2}(rn)^{-1/2}\right) \quad \text{for any sequence } s_n, \tag{3.9}$$

which follows from (3.4). In fact, (3.4) is necessary (in addition to sufficient) for the asymptotic unbiasedness and consistency of $v_{J(d)}$ as shown by the following theorem. Therefore (3.9) is also necessary for the asymptotic unbiasedness and consistency of $v_{J(d)}$.

THEOREM 3. *Assume that* (2.1) *holds.*

(i) *Condition* (3.4) *is necessary and sufficient for the asymptotic unbiasedness of* $v_{J(d)}$.

(ii) *Condition* (3.4) *is sufficient for the consistency of* $v_{J(d)}$. *It is also a necessary condition if in addition* $\{(nr/dN)\sum_s U_s^2\}$ *is uniformly integrable.*

PROOF. Sufficiency was already proved in Theorem 1.

(i) Using the same notation as in the proof of Theorem 1,

$$nEv_{J(d)} = \sigma^2 + 2\frac{nr}{dN}E\left(\sum_s L_s U_s\right) + \frac{n\tau_r}{d}.$$

Since $\Sigma_s L_s = 0$, $\Sigma_s L_s U_s = \Sigma_s L_s R_{n,s}$. Note that for any s,

$$\begin{aligned} EL_s R_{n,s} &= \frac{d}{n} E\left\{\left[\frac{1}{r}\sum_{i\in s}\phi_F(X_i) - \frac{1}{d}\sum_{i\in\bar{s}}\phi_F(X_i)\right]R_{n,s}\right\} \\ &= \frac{d}{n} E\left\{\left[\frac{1}{r}\sum_{i\in s}\phi_F(X_i)\right]R_{n,s}\right\} \\ &= \frac{d}{n} E\left\{\left[\frac{1}{r}\sum_{1}^{r}\phi_F(X_i)\right]R_r\right\} = \frac{d}{n}E[\phi_F(X_1)R_r], \end{aligned}$$

where R_r is $R_{n,s}$ with $s = \{1, \dots, r\}$. Then

$$\left|\frac{nr}{dN}E\left(\sum_s L_s U_s\right)\right| = |rE[\phi_F(X_1)R_r]| = o(1),$$

where the last equality follows from (2.1) and Lemma 1. Hence $nEv_{J(d)} = \sigma^2 + n\tau_r/d + o(1)$ and $nEv_{J(d)} = \sigma^2 + o(1)$ iff (3.4) holds.

(ii) It is easy to see that $\{(nr/dN)\Sigma_s L_s^2\}$ is uniformly integrable. From

$$nv_{J(d)} \le \frac{2nr}{dN}\sum_s\left(L_s^2 + U_s^2\right),$$

$\{nv_{J(d)}\}$ is uniformly integrable. Hence if $nv_{J(d)} = \sigma^2 + o_p(1)$, then $nEv_{J(d)} = \sigma^2 + o(1)$. Thus by part (i), (3.4) holds. □

As a final result of this section, we give some sufficient conditions to ensure that $v_{J(d)}$ and $\tilde{v}_{J(d)}$, (1.3), are asymptotically equivalent.

PROPOSITION 1. *$v_{J(d)}$ and $\tilde{v}_{J(d)}$ are asymptotically equivalent under* (i) *condition* (3.4) *or* (ii) *conditions* (2.1) *and* (3.7).

PROOF. Condition (3.4) implies $(nr/dN)\Sigma_s U_s^2 = o_p(1)$, which implies (1.4). This proves (i). (ii) follows since from Corollary 1 and Theorem 3, (2.1) and (3.7) imply (3.4). □

4. Results for $v_{J(d)}$ with bounded d. From the computational point of view, the delete-1 jackknife $v_{J(1)}$ is the most convenient one. It is clear from Theorems 1–3 that its consistency requires the point estimator $\hat{\theta}$ to be very smooth. In this section the consistency and asymptotic unbiasedness of $v_{J(d)}$ with bounded d are established. For bounded d, (3.4) becomes

$$\tau_n = o(n^{-1}). \tag{4.1}$$

Hence we have the following corollary of Theorem 3. Note that (4.1) implies that the point estimator $\hat{\theta}$ is very smooth.

COROLLARY 2. (i) *Condition* (4.1) *is necessary and sufficient for the asymptotic unbiasedness of $v_{J(d)}$ with bounded d.*

(ii) *Condition* (4.1) *is sufficient for the consistency of* $v_{J(d)}$ *with bounded* d. *If* $\{(nr/dN)\Sigma_s U_s^2\}$ *is uniformly integrable, then* (4.1) *is also necessary.*

For a smooth $\hat{\theta}$ with $ER_n^2 = O(n^{-2})$, condition (3.4) [or equivalently (4.1) for bounded d] can be verified through the following lemma, which holds for any d, bounded or unbounded, as long as $r \to \infty$. It will be used in Theorem 5 and Example 5.

LEMMA 2. *Suppose that*

$$ER_n^2 = \frac{a}{n^2} + o(n^{-2}) \quad \textit{and} \quad ER_r R_n = \frac{a}{rn} + o(r^{-2}),$$

where a is independent of n. Then (3.4) *holds.*

PROOF. From (3.2),

$$\begin{aligned}\frac{n\tau_r}{d} &= \frac{nr}{d}E(R_r - R_n)^2 = \frac{nr}{d}\left(ER_r^2 + ER_n^2 - 2ER_rR_n\right)\\ &= \frac{nr}{d}\left[\frac{a}{r^2} + \frac{a}{n^2} - 2\frac{a}{rn} + o(r^{-2})\right]\\ &= \frac{nra}{d}\left[\left(\frac{1}{r} - \frac{1}{n}\right)^2 + o(r^{-2})\right] = a\left[\frac{d}{rn} + \frac{n}{d}o(r^{-1})\right],\end{aligned}$$

which converges to zero as long as $r \to \infty$. □

Another way of defining smoothness is by the concept of differentiability. For $\hat{\theta} = T(F_n)$, where F_n is the empirical distribution of $X_1, \dots, X_n$ and T is a differentiable functional in a suitable sense, we will show that $v_{J(d)}$ with bounded d is consistent. For a strongly Fréchet differentiable T [Definition 2(i), below], Parr (1985) proved that $v_{J(1)}$ is strongly consistent, i.e., $nv_{J(1)} - \sigma^2 \to 0$ a.s. His result can be extended to any bounded d (Theorem 4). The same result holds for T which is second order Fréchet differentiable [Definition 2(ii)].

DEFINITION 2. (i) A functional T is said to be strongly Fréchet differentiable at F with respect to a norm $\|\ \|$ on Ξ iff there exists a ϕ_F: $\mathbf{R} \to \mathbf{R}$ depending only on T and F such that $E\phi_F(X_1) = 0$, $0 < E\phi_F^2(X_1) < \infty$ and

$$\frac{|T(G) - T(H) - \int\phi_F(x)d[G - H](x)|}{\|G - H\|} \to 0$$

as $\|G - F\| + \|H - F\| \to 0$, for $G, H \in \Xi$.

(ii) A functional T is said to be second order Fréchet differentiable at F with respect to $\|\ \|$ iff there are $\phi_F(x)$ and $\psi_F(x, y)$ such that $E\phi_F(X_1) = 0$, $0 < E\phi_F^2(X_1) < \infty$, $\psi_F(x, y) = \psi_F(y, x)$, $E\psi_F(x, X_1) = 0$ for any x, $E\psi_F^2(X_1, X_2) <$

∞ and

$$\frac{|T(G) - T(F) - \int\phi_F(x)\,dG(x) - \int\int\psi_F(x, y)\,dG(x)\,dG(y)|}{\|G - F\|^2} \to 0$$

$$\text{as } \|G - F\| \to 0.$$

Note that neither of the definitions implies the other. Examples of T satisfying (i) or (ii) are given in Section 6.

THEOREM 4. *If* $\hat{\theta} = T(F_n)$ *is strongly Fréchet differentiable with respect to the* sup norm$\|\ \|_\infty$, *then for any bounded* d,

$$nv_{J(d)} - \sigma^2 \to 0 \quad a.s. \tag{4.2}$$

If $\{(nr/dN)\Sigma_s U_s^2\}$ *is uniformly integrable, then* $v_{J(d)}$ *is also asymptotically unbiased.*

REMARK. Unlike the other results in this paper, the consistency in (4.2) is in the strong sense.

PROOF. By the strong Fréchet differentiability of T, for any $\varepsilon > 0$, there is a $\delta_\varepsilon > 0$ such that

$$|R_{n,s} - R_n| \le \varepsilon\|F_{n,s} - F_n\|_\infty$$

for $\|F_{n,s} - F\|_\infty + \|F_n - F\|_\infty < \delta_\varepsilon$, where $F_{n,s}$ is the empirical distribution of X_i, $i \in s$. Note that $\|F_{n,s} - F_n\|_\infty \le d/n$ and $\|F_n - F\|_\infty \to 0$ a.s. Hence there is an integer $n_\varepsilon(\omega)$, where $\omega = (X_1, X_2, \dots)$, such that

$$|R_{n,s} - R_n| \le \varepsilon\|F_{n,s} - F_n\|_\infty \le \varepsilon d/n$$

for any $s \in \mathbf{S}_{n,r}$ and $n > n_\varepsilon(\omega)$. Therefore,

$$\max_{s \in \mathbf{S}_{n,r}} (R_{n,s} - R_n)^2 = o(n^{-2}) \quad \text{a.s.},$$

which with (3.1) implies

$$W_n = \frac{nr}{dN}\sum_s U_s^2 \le \frac{nr}{d}\max_{s \in \mathbf{S}_{n,r}} (R_{n,s} - R_n)^2 = o(1) \quad \text{a.s.} \tag{4.3}$$

Then (4.2) follows from (3.6), (4.3) and the Cauchy–Schwarz inequality.

Note that under (4.3), $\{W_n\}$ is uniformly integrable iff $EW_n = o(1)$, which is equivalent to (4.1) for bounded d and implies that $v_{J(d)}$ is asymptotically unbiased (Corollary 2). □

THEOREM 5. *If T is second order Fréchet differentiable with respect to* $\|\ \|_\infty$, *then for any bounded* d,

$$nv_{J(d)} - \sigma^2 \to 0 \quad \textit{in probability}. \tag{4.4}$$

If in addition (2.5) *holds and* $E\psi_F^2(X_1, X_1) < \infty$, *then* $v_{J(d)}$ *is asymptotically unbiased.*

PROOF. From the second order Fréchet differentiability of T, there are ϕ_F and ψ_F satisfying the conditions in Definition 2(ii) and

$$\hat{\theta} = \theta + \frac{1}{n}\sum_1^n \phi_F(X_i) + \frac{1}{n^2}\sum_{i,j=1}^n \psi_F(X_i, X_j) + \Gamma_n,$$

$$\hat{\theta}_s = \theta + \frac{1}{r}\sum_{i\in s} \phi_F(X_i) + \frac{1}{r^2}\sum_{i,j\in s} \psi_F(X_i, X_j) + \Gamma_{n,s},$$

where $\max_{s\in \mathbf{S}_{n,r}}\Gamma^2_{n,s} = o_p(n^{-2})$ by Definition 2(ii) and

$$\max_{s\in\mathbf{S}_{n,r}} \|F_{n,s} - F\|_\infty \le \frac{d}{n} + \|F_n - F\|_\infty \to 0 \quad \text{a.s.}$$

Since $E\psi_F(x, X_1) = 0$, by a similar proof to that of Theorem 3 of Arvesen (1969),

$$G_n = \frac{nr}{dN}\sum_s (\Delta_{n,s} - \Delta_n)^2 = o_p(1),$$

where $\Delta_n = n^{-2}\Sigma_{i,j=1}^n\psi_F(X_i, X_j)$ and $\Delta_{n,s} = r^{-2}\Sigma_{i,j\in s}\psi_F(X_i, X_j)$. Hence (4.4) holds since

$$\frac{nr}{dN}\sum_s U_s^2 \le 2\left(G_n + \frac{nr}{d}\max_{s\in\mathbf{S}_{n,r}}\Gamma^2_{n,s}\right). \tag{4.5}$$

If (2.5) holds, then following the same argument as in the proof of Lemma 1, we have

$$E\Gamma_n^2 = o(n^{-2}). \tag{4.6}$$

Decompose

$$\frac{1}{n^2}\sum_{i,j=1}^n \psi_F(X_i, X_j) = \frac{\alpha}{n} + V_n + \frac{n-1}{n}U_n,$$

where $\alpha = E\psi_F^2(X_1, X_1)$, $V_n = n^{-2}\Sigma_1^n[\psi_F(X_i, X_i) - \alpha]$ and $U_n = n^{-1}(n-1)^{-1}\Sigma_{i\ne j}\psi_F(X_i, X_j)$. Since U_n is a U-statistic and $E\psi_F(x, X_1) = 0$,

$$EU_n^2 = \frac{\rho}{n^2} + O(n^{-3}) \quad \text{and} \quad EU_rU_n = \frac{\rho}{rn} + O(n^{-3}),$$

where $\rho = 2E\psi_F^2(X_1, X_2)$. From $EV_n^2 = O(n^{-3})$ and (4.6), we have

$$ER_n^2 = \frac{\alpha+\rho}{n^2} + o(n^{-2}) \quad \text{and} \quad ER_rR_n = \frac{\alpha+\rho}{rn} + o(r^{-2}),$$

where $R_n = n^{-2}\Sigma_{i,j=1}^n\psi_F(X_i, X_j) + \Gamma_n$. Hence from Lemma 2, (4.1) holds and $v_{J(d)}$ with bounded d is asymptotically unbiased. □

From (4.3) and (4.5) in the proofs of Theorems 4 and 5, either strongly Fréchet differentiability or second order Fréchet differentiability implies $(nr/dN)\Sigma_sU_s^2 = o_p(1)$ and therefore (1.4). Hence we have the following corollary.

COROLLARY 3. *If T is either strongly Fréchet differentiable or second order Fréchet differentiable with respect to $\| \|_\infty$, then $v_{J(d)}$ and $\tilde{v}_{J(d)}$ are asymptotically equivalent.*

5. Extension to balanced subsampling. The computation of $v_{J(d)}$ requires $\binom{n}{d}$ evaluations of $\hat{\theta}_s$. The computational complexity increases rapidly as d increases. To circumvent this, $\hat{\theta}_s$ may be evaluated only for a collection of subsets chosen from $\mathbf{S}_{n,r}$ in a systematic manner. We will show in this section that all the results of Section 3 hold for one such scheme called "balanced subsampling."

Let $\mathbf{B} = \{s_1, \dots, s_b\}$ be a collection of b subsets of size r satisfying the following two properties:

(1) Every i, $1 \le i \le n$, appears in the same number, denoted by f, of subsets in $\mathbf{B}$.
(2) Every pair (i, j), $1 \le i < j \le n$, appears together in the same number, denoted by λ, of subsets in $\mathbf{B}$.

If each subset is treated as a "block" and each i as a "treatment," $\mathbf{B}$ is a balanced incomplete block design (BIBD). Standard results for BIBD [John (1971)] give the relations

$$br = nf, \tag{5.1}$$

$$(r-1)f = \lambda(n-1), \tag{5.2}$$

$$b \ge n. \tag{5.3}$$

The relation (5.3) is called Fisher's inequality in the BIBD literature. It implies that the number of computations in balanced subsampling is at least n. If $b = n$, such a plan is called a minimal balanced subsampling. The corresponding design is called symmetric BIBD [John (1971)].

In a balanced subsampling scheme, $\hat{\theta}_s$ is evaluated only for those s in $\mathbf{B}$. Extensions for $v_{J(d)}$ and $\tilde{v}_{J(d)}$ are, respectively,

$$v_{BJ(d)} = \frac{r^2}{n(n-1)(f-\lambda)} \sum_{s \in \mathbf{B}} \left(\hat{\theta}_s - \hat{\theta}\right)^2$$

and

$$\tilde{v}_{BJ(d)} = \frac{r^2}{n(n-1)(f-\lambda)} \sum_{s \in \mathbf{B}} \left(\hat{\theta}_s - \frac{1}{b} \sum_{s \in \mathbf{B}} \hat{\theta}_s\right)^2.$$

They include $v_{J(d)}$ and $\tilde{v}_{J(d)}$ as special cases since it is easy to show $v_{BJ(d)} = v_{J(d)}$ and $\tilde{v}_{BJ(d)} = \tilde{v}_{J(d)}$ by taking $b = \binom{n}{d}$, $f = \binom{n-1}{d}$ and $\lambda = \binom{n-2}{d}$. This scheme was proposed by Mellor (1972) in another context.

When $\hat{\theta}$ is linear in the sense of (1.5), it can be shown by straightforward algebra, (5.1) and (5.2) that

$$v_{BJ(d)} = \tilde{v}_{BJ(d)} = v_{J(1)} = \frac{1}{n(n-1)} \sum_1^n \left(\phi_F(X_i) - \frac{1}{n} \sum_1^n \phi_F(X_i)\right)^2. \tag{5.4}$$

We now show that Theorems 1 and 3 for $v_{J(d)}$ (and therefore their consequences) hold for $v_{BJ(d)}$ for *any* choice of $\mathbf{B}$ defined above.

An analog of Theorem 1 is stated as follows.

THEOREM 1A. *Under either* (3.3) *or* (3.4), $v_{BJ(d)}$ *is consistent and asymptotically unbiased.*

PROOF. Consider the decomposition

$$nv_{BJ(d)} = c \sum_{s \in \mathbf{B}} (L_s + U_s)^2, \qquad c = \frac{r^2}{(n-1)(f-\lambda)}, \tag{5.5}$$

where L_s and U_s are defined in (3.5) and (3.1), respectively. From the definition of BIBD, $c\Sigma_{s \in \mathbf{B}} L_s^2$ is identical to $nv_{BJ(d)}$ for linear $\hat{\theta}$, which according to (5.4) has mean σ^2 and converges to σ^2 a.s. From the exchangeability of U_s in s and (5.1) and (5.2),

$$E\left(c \sum_{s \in \mathbf{B}} U_s^2\right) = \frac{rb}{(n-1)(f-\lambda)} rEU_s^2 = \frac{n}{d}\tau_r.$$

The rest of the proof is the same as that of Theorem 1. □

An analog of Theorem 3 is the following.

THEOREM 3A. *Assume that* (2.1) *holds.*

(i) *Condition* (3.4) *is necessary and sufficient for the asymptotic unbiasedness of* $v_{BJ(d)}$.

(ii) *Condition* (3.4) *is sufficient for the consistency of* $v_{BJ(d)}$. *It is also a necessary condition if in addition* $\{c\Sigma_{s \in \mathbf{B}} U_s^2\}$ *is uniformly integrable, where c is defined in* (5.5).

PROOF. It is easy to show that

$$nEv_{BJ(d)} = \sigma^2 + cE\left(\sum_{s \in \mathbf{B}} L_s U_s\right) + \frac{n\tau_r}{d}.$$

Following the proof of Theorem 3(i), the proof of (i) is completed by noting that

$$\sum_{s \in \mathbf{B}} L_s = 0, \qquad \sum_{s \in \mathbf{B}} L_s U_s = \sum_{s \in \mathbf{B}} L_s R_{n,s}$$

and

$$cE\left(\sum_{s \in \mathbf{B}} L_s U_s\right) = cbE(L_s U_s) = \frac{cbd}{n} E(\phi_F(X_1) R_r) = rE(\phi_F(X_1) R_r) = o(1).$$

The proof of (ii) is very similar to that of Theorem 3(ii). □

The results in Theorems 1A and 3A also hold for $\tilde{v}_{BJ(d)}$ with the minor change that U_s is replaced by $\tilde{U}_s = R_{n,s} - b^{-1}\Sigma_{s \in \mathbf{B}} R_{n,s}$ and (3.4) is replaced by

$\lim_{n\to\infty} d^{-1}n\tilde{\tau}_r = 0$, $\tilde{\tau}_r = rE\tilde{U}_s^2$. From (5.1) and (5.2),

$$v_{BJ(d)} = \tilde{v}_{BJ(d)} + \frac{r}{d}\left(\frac{1}{b}\sum_{s\in\mathbf{B}}\hat{\theta}_s - \hat{\theta}\right)^2.$$

Hence we have the following analog of Proposition 1. The proof is very similar and therefore omitted.

PROPOSITION 1A. (i) *Under* (3.4), $v_{BJ(d)}$ *and* $\tilde{v}_{BJ(d)}$ *are asymptotically equivalent, i.e.*,

$$v_{BJ(d)} = \tilde{v}_{BJ(d)} + o_p(n^{-1}). \tag{5.6}$$

(ii) *Under* (2.1) *and* (3.7), (5.6) *holds.*

6. Some examples. Using the general results in Sections 3 and 4, we discuss the consistency and asymptotic unbiasedness of $v_{J(d)}$ in several situations.

EXAMPLE 1. Sample quantiles. Let $\theta = T(F) = \inf\{x\colon F(x) \geq p\}$ be the p-quantile of F, $0 < p < 1$. Suppose that the distribution F has a positive first derivative at θ, i.e., $F'(\theta) > 0$. Then by Ghosh (1971), the sample quantile $\hat{\theta} = T(F_n)$ has the expansion

$$\hat{\theta} = \theta + \frac{1}{n}\sum_1^n \phi_F(X_i) + R_n, \qquad R_n = o_p(n^{-1/2}), \tag{6.1}$$

where $\phi_F(x) = (p - I(x \leq \theta))/F'(\theta)$ and $I(A)$ is the indicator function of the set A. Here

$$\sigma^2 = E\phi_F^2(X) = p(1-p)/(F'(\theta))^2.$$

In Section 1, $v_{J(1)}$ for the sample median was shown to be inconsistent. To show that $v_{J(1)}$ is also asymptotically biased, first note that under (2.5), $n\operatorname{Var}\hat{\theta} \to \sigma^2$. On the other hand,

$$\liminf_{n\to\infty} E(nv_{J(1)}) \geq \sigma^2 E(\chi_2^2/2)^2 = 2\sigma^2 > \sigma^2,$$

by following essentially the proof of Lemma 5.1.2 of Lehmann (1983). Inconsistency and asymptotic biasedness for general p and bounded d can be shown by a similar argument.

Theorem 3 can be used to explain the inconsistency of $v_{J(1)}$. Let

$$R_{ni} = \hat{\theta}_{-i} - \theta - \frac{1}{n-1}\sum_{j\neq i}\phi_F(X_j).$$

Then for $n = 2m$ and $i = 1,\dots,m$,

$$\begin{aligned} n(R_{ni} - R_n) &= n\left[\hat{\theta}_{-i} - \hat{\theta} + \frac{1}{n}\phi_F(X_i) - \frac{1}{n(n-1)}\sum_{j\neq i}\phi_F(X_j)\right] \\ &= n\frac{X_{(m+1)} - X_{(m)}}{2} + \frac{1/2 - I(X_i \leq \theta)}{F'(\theta)} + O_p(n^{-1/2}). \end{aligned}$$

Since $(\frac{1}{2} - I(X_i \le \theta))/F'(\theta)$ is bounded and discrete and $n(X_{(m+1)} - X_{(m)})/2$ converges in distribution to a random variable with a continuous distribution (see Section 1), their sum does not converge to zero in probability. Thus (3.9) is not satisfied. Hence $v_{J(1)}$ is not asymptotically unbiased or consistent.

The previous argument shows that, in order for $v_{J(d)}$ to be asymptotically unbiased and consistent, d has to be unbounded. For d satisfying condition (3.7), $v_{J(d)}$ is asymptotically unbiased and consistent under conditions (2.2) and (2.5) (Corollary 1 and Lemma 1). From (6.1), (2.2) is satisfied. Hence we only need to verify (2.5). This can be done under the weak condition

(6.2) $\quad E|X_1|^{\delta} < \infty$ for some $\delta > 0$ and F'' exists in a neighborhood of θ.

In fact, under (6.2), the asymptotic unbiasedness and consistency of $v_{J(d)}$ can be achieved with a smaller d, since Duttweiler (1973) proved that

$$ER_n^2 = \left[\frac{2p(1-p)}{\pi F'(\theta)}\right]^2 n^{-3/2} + o(n^{-7/4+\varepsilon}).$$

Hence if $d = d_n$ satisfies $n^{1/2}/d \to 0$ and $r \to \infty$, then (3.3) is satisfied and $v_{J(d)}$ is consistent and asymptotically unbiased from Theorem 1. A smaller d will suffice if τ_r in (3.2) can be shown to have an order smaller than that of δ_r, which in this case is $O(r^{-1/2})$.

Without condition (2.5) or (6.2), the consistency of $v_{J(d)}$ with d satisfying $n^{1/2}/d \to 0$ can be established by a different approach. See Shao (1988).

Example 2. *L*-estimators. Let $X_{(1)}, \dots, X_{(n)}$ be the order statistics of $X_1, \dots, X_n$. Consider *L*-estimators of the form

$$\hat{\theta} = \sum_1^n c_{ni} X_{(i)}.$$

Parr and Schucany (1982) and Parr (1985) proved that for $\hat{\theta}$ with "smooth" c_{ni}, $v_{J(1)}$ is consistent. On the other hand, for $\hat{\theta}$ with discontinuous c_{ni} such as the sample quantiles, $v_{J(1)}$ is inconsistent as shown in Example 1. It was not known whether $v_{J(1)}$ works well in situations between these two extremes. In the following we will show that Corollary 1 can be used to handle a large class of not very smooth *L*-estimators.

Most *L*-estimators can be written as $T(F_n)$, where $T(F) = T_1(F) + T_2(F)$ with

$$T_1(F) = \int xJ(F(x))\, dF(x) \quad \text{and} \quad T_2(F) = \sum_1^k a_j F^{-1}(p_j),$$

where $J(t)$ is a function on $(0, 1)$, k is fixed, $0 < p_j < 1$ and $F^{-1}(t) = \inf\{x\colon F(x) \ge t\}$. Note that $T_2(F_n)$ is a finite combination of sample quantiles. Since the sample quantiles were studied in Example 1, we need only consider $T_1(F_n)$.

Parr (1985) showed that, if $J(t)$ is bounded, continuous a.e. Lebesgue and a.e. F^{-1}, and zero outside a compact subset of $(0, 1)$, then T_1 is strongly Fréchet differentiable with respect to the sup norm$\| \|_\infty$. Hence $v_{J(d)}$ with bounded d is consistent according to Theorem 4. If the derivative of J exists and is smooth enough, T_1 is second order Fréchet differentiable (see Example 5) and therefore Theorem 5 applies. If neither the support of J is compact nor $J'(t)$ is smooth, the behavior of $v_{J(d)}$ with bounded d remains unknown. However, the application of our results for unbounded d requires weaker conditions on J. This is stated and proved in the following proposition.

PROPOSITION 2. *Suppose that* $\hat{\theta} = T(F_n)$, $T(F) = \int xJ(F(x))\,dx$ *and*

$$0 < \sigma^2 = \int\int J(F(x))J(F(y))[F(\min(x, y)) - F(x)F(y)]\,dx\,dy < \infty.$$

(i) *If J is bounded and continuous a.e. F^{-1} and $EX_1^2 < \infty$, then $v_{J(d)}$ with d satisfying* (3.7) *is consistent and asymptotically unbiased.*

(ii) *Assume that J is Lipschitz-continuous on* $(0,1)$ *and*

$$\int F(x)[1 - F(x)]\,dx < \infty. \tag{6.3}$$

Then $v_{J(d)}$ is consistent and asymptotically unbiased if both r and $d \to \infty$.

PROOF. In all cases, $\hat{\theta}$ has the expansion (1.7) with

$$\phi_F(x) = \int [F(y) - I(y \geq x)]\,J(F(y))\,dy$$

and

$$R_n = \int W_{F_n, F}(x)[F(x) - F_n(x)]\,dx,$$

where $W_{G,F}(x) = [G(x) - F(x)]^{-1}[K(G(x)) - K(F(x))] - J(F(x))$, $K(t) = \int_0^t J(u)\,du$ if $G(x) \neq F(x)$ and $W_{G,F}(x) = 0$ if $G(x) = F(x)$.

For (i), (2.2) and (2.5) follow from Theorem 1 of Stigler (1974). Hence by Lemma 1, (2.1) holds and the result follows from Corollary 1.

For (ii), since J is Lipschitz-continuous, $|W_{G,F}(x)| \leq c|G(x) - F(x)|$ for a constant $c > 0$. Therefore

$$ER_n^2 \leq c^2 \int\int E\{[F_n(x) - F(x)]^2[F_n(y) - F(y)]^2\}\,dx\,dy = O(n^{-2})$$

under (6.3). Then Theorem 1 applies since (3.3) holds when both r and $d \to \infty$. □

Note that part (ii) ensures the consistency of $v_{J(d)}$ for $d \to \infty$ at any rate if J is smoother. The conclusions of Proposition 2 may hold under other conditions on J and F given in Stigler (1974), Boos (1979), Mason (1981), etc.

EXAMPLE 3. M-estimators. We consider the M-functional $\theta = T(F)$, which is a solution of

$$\int \psi(x, \theta)\, dF(x) = 0, \tag{6.4}$$

and the corresponding M-estimator $\hat{\theta} = T(F_n)$, where ψ is a function from $\mathbf{R}^2$ to $\mathbf{R}$. Here we assume that θ is a unique solution of (6.4) and $0 < \int \psi^2(x, \theta)\, dF(x) < \infty$. Examples of these estimators can be found in Serfling (1980), Chapter 7.

Reeds (1978) proved the consistency of $v_{J(1)}$ by imposing some conditions on ψ and F (see his Assumptions L.1–L.5). Some of his conditions are rather restrictive since they fail to cover some commonly used ψ functions such as Huber's (1964), Hampel's (1974) and that corresponding to the least pth power estimate [Serfling (1980) Chapter 7]. Details can be found in Shao and Wu (1986).

It is generally unknown whether $v_{J(1)}$ is consistent. For $v_{J(d)}$ with d satisfying (3.7), its consistency and asymptotic unbiasedness follow from applying Corollary 1, that is, to check conditions (2.2) and (2.5) for the remainder term R_n of the expansion,

$$\hat{\theta} = \theta + \frac{1}{n}\sum_1^n \phi(X_i, \theta) + R_n, \qquad \phi(x, \theta) = \psi(x, \theta)/I(\theta)$$

and

$$I(\theta) = -E\left[\frac{\partial}{\partial\theta}\psi(X_1, \theta)\right].$$

Condition (2.2) was shown to hold under various regularity conditions on ψ and F in Shao and Wu (1986). The verification of (2.5) for general ψ is more difficult. For some specific ψ functions, (2.5) can be verified. For example, when $\psi = f_\theta'(x)/f_\theta(x)$, where $f_\theta(x)$ is the density function of X_1, both (2.2) and (2.5) hold [Ibragimov and Khas'minskii (1972/1973)] under some regularity conditions on $f_\theta(x)$.

Clarke (1983, 1986) proved the Fréchet differentiability of $T(F)$ in a variety of situations. Since strong Fréchet differentiability can sometimes be obtained from Fréchet differentiability [Parr (1985)], Clarke's results may be useful in view of Theorem 4.

EXAMPLE 4. Bayes estimators. Suppose $f_\theta(x)$ is the density function of X_1 and $\pi(\theta)$ is a prior of θ. Ibragimov and Khas'minskii (1972/1973) proved that the Bayes estimator $\hat{\theta}_B$ satisfies (2.2) and (2.5) under their conditions 1–4. Hence Corollary 1 can be used to obtain a consistent and asymptotically unbiased variance estimator of $\hat{\theta}_B$.

EXAMPLE 5. The case of bounded d.

(a) U-statistics. A U-statistic is defined to be

$$\hat{\theta} = \binom{n}{k}^{-1} \sum_k h(X_{i1}, \dots, X_{ik}),$$

where $\sum_k$ denotes summation over the $\binom{n}{k}$ combinations of k distinct elements $\{i_1, \dots, i_k\}$ from $\{1, \dots, n\}$, k is fixed and $h = h(x_1, \dots, x_k)$ is a symmetric kernel satisfying $Eh^2(X_1, \ldots, X_k) < \infty$. $\hat{\theta}$ is an unbiased estimator of $\theta = Eh(X_1, \ldots, X_k)$. From Hoeffding (1948), $\hat{\theta}$ has the expansion (1.7) with

$$\phi_F(x) = kE\left[h(X_1, \ldots, X_k) - \theta | X_1 = x\right] \quad \text{and} \quad ER_n^2 = \frac{k^2(k-1)^2\zeta}{n^2} + O(n^{-3})$$

for some constant $\zeta > 0$. Here we only consider $\hat{\theta}$ with $E\phi_F^2(X_1) > 0$. A straightforward calculation shows that

$$ER_rR_n = \frac{k^2(k-1)^2\zeta}{rn} + O(n^{-3})$$

for any $r = n - d$ with bounded d. Hence from Lemma 2, (4.1) is satisfied and $v_{J(d)}$ with bounded d is asymptotically unbiased and consistent.

Consistency for the special case of $v_{J(1)}$ was established by Arvesen (1969).

(b) Strongly Fréchet differentiable function. Several classes of strongly Fréchet differentiable functionals are given in Parr (1985), Appendices (a)–(d). From Theorem 4, $v_{J(d)}$ with bounded d is strongly consistent.

(c) Second order Fréchet differentiable functional. Consider the L-functional $T(F) = \int xJ(F(x))\,dx$ with continuously differentiable $J(t)$, $t \in [0,1]$, where F satisfies (6.3) and

$$\phi_F(x) = \int [F(u) - I(u \geq x)]\, J(F(u))\, du$$

and

$$\psi_F(x, y) = -\int [F(u) - I(x \geq u)][F(u) - I(y \geq u)]\, J'(F(u))\, du$$

satisfy $E\phi_F^2(X_1) > 0$,

$$E\phi_F^2(X_1) < \infty \quad \text{and} \quad E\psi_F^2(X_1, X_2) < \infty. \tag{6.5}$$

A sufficient condition for (6.3) and (6.5) is $EX_1^2 < \infty$ [Serfling (1980), page 288]. A similar argument to the proof of Proposition 2(ii) shows that T is second order Fréchet differentiable. If in addition, $E\psi_F^2(X_1, X_1) < \infty$, then (2.5) holds. Hence the consistency and asymptotic unbiasedness of $v_{J(d)}$ with bounded d follow from Theorem 5.

7. Concluding remarks.

Choice of subset size. For sufficiently smooth $\hat{\theta}$, the delete-1 jackknife $v_{J(1)}$ is recommended because it has desirable asymptotic properties (Corollary 2) and is computationally simpler. For nonsmooth $\hat{\theta}$, $v_{J(d)}$ with $d = \lambda n$, $0 < \lambda < 1$, is recommended because it has desirable asymptotic properties as long as condition (2.1) is satisfied (Corollary 1). The jackknife histogram with this choice of d also has desirable asymptotic properties [Wu (1987)]. If there exists a balanced subsampling plan with the total number of balanced subsets much smaller than $\binom{n}{d}$, $v_{BJ(d)}$ will be recommended because it shares the same asymptotic properties of $v_{J(d)}$ and is computationally less intensive. On the other hand, a separate software for enumerating the balanced subsets is required. If an economic balanced subsampling plan is not available, or for other computational considerations, random sampling of subsets may be used. A corresponding theory can be found in Shao (1989).

The condition (2.5): $\operatorname{Var}\hat{\theta} = \sigma^2/n + o(n^{-1})$. This condition requires $\operatorname{Var}\hat{\theta} < \infty$, thus excluding some common estimators such as the ratio of means. For the purpose of inference, one is often satisfied with estimating σ^2/n, the variance of the limiting distribution of $\hat{\theta}$. It is not clear whether our approach can be modified to handle this situation. Even when $\operatorname{Var}\hat{\theta} < \infty$, verification of (2.5) can be technically difficult. However, for some estimators such as smooth functions of sample means (including the ratio of means), the sample quantiles and M-estimators with some ψ functions, by adopting different approaches, the consistency of $v_{J(d)}$ can also be established without checking (2.5). This will be reported later.

Consistency of the grouped jackknife variance estimator. This method works by dividing n into g groups each of size h, $n = gh$, and computing $\hat{\theta}_{-i}$ with the ith group of observations removed from the sample. The grouped jackknife variance estimator is

$$v_{gJ(h)} = \frac{g-1}{g}\sum_1^g \left(\hat{\theta}_{-i} - \hat{\theta}\right)^2.$$

It is obvious that, for the consistency of $v_{gJ(h)}$, g should go to infinity as $n \to \infty$. If the group size h, which plays the same role as d in our $v_{J(d)}$, remains bounded and $g \to \infty$, it can be shown that $v_{gJ(h)}$ behaves asymptotically like $v_{J(h)}$. Therefore for nonsmooth $\hat{\theta}$ such as the sample quantiles, it is necessary for the consistency of $v_{gJ(h)}$ that both g and $h \to \infty$, resulting in a more stringent requirement on the sample size $n = gh$. Computational saving seems to be the main reason for adopting this method. The same can be achieved by balanced subsampling, which has better asymptotic properties.

Extension to the multivariate case. For vector $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_p)$, $v_{J(d)}$ and $v_{BJ(d)}$ have the obvious extension with $\hat{\theta}_s$ and $\hat{\theta}$ meaning vectors instead of scalars. The proofs of asymptotic unbiasedness and consistency for each variance or covariance component of $v_{J(d)}$ are the same as before.

Acknowledgments. We would like to thank the referees and the Associate Editor for their helpful comments. A referee's suggestion led to our Proposition 2(ii).

REFERENCES

ARVESEN, J. N. (1969). Jackknifing U-statistics. *Ann. Math. Statist.* **40** 2076–2100.

BHARGAVA, R. P. (1983). A property of the jackknife estimation of the variance when more than one observation is omitted. *Sankhyā Ser. A* **45** 112–119.

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

CLARKE, B. R. (1983). Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations. *Ann. Statist.* **11** 1196–1205.

CLARKE, B. R. (1986). Nonsmooth analysis and Fréchet differentiability of M-functionals. *Probab. Theory Related Fields* **73** 197–209.

DUTTWEILER, D. L. (1973). The mean-square error of Bahadur's order-statistic approximation. *Ann. Statist.* **1** 446–453.

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

GHOSH, J. K. (1971). A new proof of the Bahadur representation of quantiles and an application. *Ann. Math. Statist.* **42** 1957–1961.

GHOSH, M., PARR, W. C., SINGH, K. and BABU, G. J. (1984). A note on bootstrapping the sample median. *Ann. Statist.* **12** 1130–1135.

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–397.

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

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

IBRAGIMOV, I. A. and KHAS'MINSKII, R. Z. (1972 / 1973). Asymptotic behavior of some statistical estimators. I, II. *Theory Probab. Appl.* **17** 445–462; **18** 76–91.

JOHN, P. W. M. (1971). *Statistical Design and Analysis of Experiments*. Macmillan, New York.

LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.

MASON, D. M. (1981). Asymptotic normality of linear combinations of order statistics with a smooth score function. *Ann. Statist.* **9** 899–904.

MELLOR, R. W. (1972). Subsample replication variance estimators. Ph.D. dissertation, Harvard Univ.

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

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

PARR, W. C. and SCHUCANY, W. R. (1982). Jackknifing L-statistics with smooth weight functions. *J. Amer. Statist. Assoc.* **77** 629–638.

RAO, J. N. K. and WU, C. F. J. (1985). Inference from stratified samples: Second order analysis of three methods for nonlinear statistics. *J. Amer. Statist. Assoc.* **80** 620–630.

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

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

SHAO, J. (1988). Consistency of jackknife estimators of the variances of sample quantiles. *Comm. Statist. A–Theory Methods* **17** 3017–3028.

SHAO, J. (1989). The efficiency and consistency of approximations to the jackknife variance estimators. *J. Amer. Statist. Assoc.* **84** 114–119.

SHAO, J. and WU, C. F. J. (1986). Some general theory for the jackknife. Technical Report 797, Dept. Statistics, Univ. Wisconsin, Madison.

STIGLER, S. M. (1974). Linear functions of order statistics with smooth weight functions. *Ann. Statist.* **2** 676–693.

WU, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.* **14** 1261–1350.

WU, C. F. J. (1987). On the asymptotic properties of the jackknife histograms. Technical Report, Dept. Statistics, Univ. Wisconsin, Madison.

DEPARTMENT OF STATISTICS
MATHEMATICAL SCIENCES BUILDING
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907

DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE
UNIVERSITY OF WATERLOO
WATERLOO, ONTARIO
CANADA N2L 3G1