The Annals of Statistics
1994, Vol. 22, No. 2, 946–967

L-STATISTICS IN COMPLEX SURVEY PROBLEMS[1]

BY JUN SHAO

University of Wisconsin–Madison

We study linear combinations of order statistics (*L*-statistics) in survey problems with a stratified multistage sampling design. Two general types of *L*-statistics are considered: smooth *L*-statistics with weights generated by a smooth function and nonsmooth *L*-statistics (sample quantiles). The trimmed sample mean, the decile mean and variance, the sample Lorenz curve and the sample Gini's family parameters are examples of smooth *L*-statistics or functions of smooth *L*-statistics used in survey problems. It is shown that under weak conditions the smooth *L*-statistics are asymptotically normal and their asymptotic variances can be consistently estimated by jackknifing. For the sample quantiles, their asymptotic normality requires more conditions on the finite population distribution functions. Consistent estimators for the asymptotic variances of the sample quantiles are derived. Asymptotic validity of the Woodruff's confidence intervals for population quantiles is also proved.

1. Introduction. One of the most commonly used class of statistics in statistical analysis is the class of linear combinations of order statistics, termed *L*-statistics in the literature. Simple statistics such as the sample mean, the sample quantiles and the trimmed sample mean are special types of *L*-statistics. Asymptotic properties of *L*-statistics are well known in the case where the data are independent and identically distributed (i.i.d.) observations [e.g., Serfling (1980), Chapter 8]. Theoretical studies of the *L*-statistics in some non-i.i.d. cases have received considerable attention in recent years. For example, Bickel (1973), Koenker and Bassett (1978), Ruppert and Carroll (1980), Koenker and Portnoy (1987) and Welsh (1987) established many asymptotic properties of *L*-statistics in linear regression problems.

The purpose of this work is to study asymptotic properties of *L*-statistics with complex survey data. Sample surveys often use one or a combination of several of the following sampling methods: stratified sampling, cluster sampling, unequal probability sampling and multistage sampling [Kish and Frankel (1974)]. Simple random sampling is rarely used in practice because of both practical and theoretical considerations. Due to the complexity of the sampling design, the resulting data are heavily non-i.i.d.

Throughout the paper we consider the following commonly used stratified multistage sampling design. The population under consideration has been stratified into L strata. From each stratum, some first-stage units (clusters)

Received August 1991; revised March 1993.

[1]The research was supported by the Natural Sciences and Engineering Research Council of Canada.

AMS 1991 *subject classifications.* Primary 62D05; secondary 62G20, 62G09.

Key words and phrases. Asymptotic normality, finite population, multistage sampling, trimmed mean, Lorenz curve, Gini's family, quantile, variance estimation, jackknife.

are selected using unequal probability sampling with replacement (or simple random sampling without replacement) and the samples are selected independently across the strata. Within each first-stage unit (cluster), some ultimate units are sampled according to some sampling methods. Note that we do not need to specify the number of stages and the sampling methods used after the first-stage sampling.

Under this stratified multistage sampling design, Krewski and Rao (1981) established some asymptotic results for the sample mean, a special L-statistic and functions of several sample means (e.g., ratio estimators). The use of other L-statistics in complex survey problems has become more and more popular in recent years. For example, the sample quantiles are important statistics for survey problems related to earnings. An asymptotic study is given by Francisco and Fuller (1991). Other examples, including the trimmed means, the Gini's family and Lorenz curve, are described in Section 2.

Section 2 contains an introduction of notation, a formal definition of L-statistics and some examples. The asymptotic normality of the L-statistics in stratified multistage survey problems is established in Sections 3 and 4. Consistent estimators of the asymptotic variances of the L-statistics are also derived. These results are useful for large-sample inferences.

2. Definitions and examples. Let $\{\mathcal{P}_k, k = 1, 2, \ldots\}$ be a sequence of finite populations. Throughout the paper k is used as the index of the finite population, but it may be omitted frequently for simplicity. Each $\mathcal{P}_k$ contains L strata, and the hth stratum contains N_h clusters. Associated with the jth ultimate unit in the ith cluster of stratum h is a characteristic $Y_{hij}, j = 1, \ldots, N_{hi}$; $i = 1, \ldots, N_h$; $h = 1, \ldots, L$. Here N_{hi} is the number of ultimate units in the ith cluster of stratum h. Note that L, N_h, N_{hi}, Y_{hij}, and so on depend on k also but the subscript k is omitted. Let $I(A)$ be the indicator function of the set A. The kth finite population distribution function is

$$F_k(x) = \frac{1}{M} \sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{j=1}^{N_{hi}} I(Y_{hij} \le x), \tag{2.1}$$

where $M = \sum_{h=1}^{L} \sum_{i=1}^{N_h} N_{hi}$.

Suppose that in the first-stage sampling, $n_h \ge 2$ clusters are selected from stratum h (independently across the strata) with probabilities $p_{hi} > 0, i = 1, \ldots, N_h; h = 1, \ldots, L; \sum_{i=1}^{N_h} p_{hi} = 1$. For a fixed h, if $p_{hi} \equiv 1/N_h$, the sampling method is simple random sampling (equal probability sampling) and the clusters are selected either with replacement or without replacement. If unequal probability sampling is used, we assume that the clusters are selected independently (with replacement). Within each cluster, some second-stage, third-stage,... units are sampled according to some methods (e.g., stratified sampling, cluster sampling, unequal probability sampling). Let n_{hi} be the number of ultimate units drawn from the ith cluster of stratum h, y_{hij} and $\tilde{w}_{hij}$ be the observed characteristic and the survey weight associated with the (h, i, j)th ultimate unit,

$j = 1, \ldots, n_{hi}$; $i = 1, \ldots, n_h$; $h = 1, \ldots, L$. Again, n_h, n_{hi}, y_{hij} and $\tilde{w}_{hij}$ depend on k but the subscript k is omitted. Suppose that the survey weights are constructed so that

$$G_k(x) = \frac{1}{M} \sum_{h=1}^{L} \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} \tilde{w}_{hij} I(y_{hij} \le x) \tag{2.2}$$

is unbiased for $F_k(x)$, that is,

$$E[G_k(x)] = F_k(x) \quad \text{for any } x$$

[see, e.g., Krewski and Rao (1981)]. However, G_k may not be a distribution function since $G_k(\infty) = M^{-1} \sum_{h=1}^{L} \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} \tilde{w}_{hij}$ is not necessarily equal to 1. Furthermore, in many cases M, the total number of ultimate units in the population, is unknown and therefore G_k is not an estimator. Hence we consider the normalized G_k:

$$\hat{F}_k(x) = G_k(x)/q_k, \tag{2.3}$$

where $q_k = G_k(\infty)$. Then $\hat{F}_k$ is a distribution function for any k and can be used to estimate F_k since the unknown M is cancelled in the ratio in (2.3).

Let $n = \sum_{h=1}^{L} n_h$ be the number of sampled clusters in the first-stage sampling. We always assume that $n \to \infty$ as $k \to \infty$. Denote $\tilde{w}_{hij}/M$ by w_{hij}. Assuming

$$\max_{j \le N_{hi}, i \le N_h, h \le L} n_{hi} w_{hij} = O\left(\frac{1}{n}\right) \tag{A1}$$

and using

$$\begin{aligned} \operatorname{Var}[G_k(x)] &\le \sum_{h=1}^{L} \sum_{i=1}^{n_h} E\left(\sum_{j=1}^{n_{hi}} w_{hij} I(y_{hij} \le x) \right)^2 \\ &\le E\left(\sum_{h=1}^{L} \sum_{i=1}^{n_h} n_{hi} \sum_{j=1}^{n_{hi}} w_{hij}^2 \right) \le \max_{j,i,h} n_{hi} w_{hij}, \end{aligned}$$

we obtain that, for any x, $G_k(x) - F_k(x) \to_p 0$. In particular, since $q_k = G_k(\infty)$ and $F_k(\infty) = 1$, $q_k - 1 \to_p 0$ and therefore

$$\hat{F}_k(x) - F_k(x) \to_p 0 \quad \text{for any } x.$$

Assumptions (A1) means that no survey weight is disproportionately large. It is of interest to see what (A1) reduces to in some simple special cases. If the sampling design is only one-stage and simple random sampling is used, then $w_{hij} = N_h/Mn_h$ and (A1) is the same as

$$\max_{h \le L} \frac{N_h}{Mn_h} = O\left(\frac{1}{n}\right), \tag{2.4}$$

an assumption given in Rao and Wu (1985). If the sampling design is a stratified two-stage sampling design and the second-stage sampling is also simple random sampling (within each cluster), then $w_{hij} = N_h N_{hi}/Mn_h n_{hi}$ and (A1) reduces to

$$\max_{i \le N_h, h \le L} \frac{N_h N_{hi}}{Mn_h} = O\left(\frac{1}{n}\right). \tag{2.5}$$

Note that N_{hi} is the number of units in the ith cluster of stratum h. If N_{hi}, $i = 1, \ldots, N_h$, are relatively the same or they are bounded, then (2.5) is the same as (2.4).

Let J be a function on the interval $[0, 1]$ and

$$T(F) = \int xJ\big(F(x)\big)\, dF(x), \tag{2.6}$$

for any distribution function F. An L-statistic is generally of the form

$$T_k = T(\widehat{F}_k) + \sum_{t=1}^{m} a_t \widehat{F}_k^{-1}(p_t), \tag{2.7}$$

where m is a fixed integer, p_t and a_t are constants, $0 < p_t < 1$ and $F^{-1}(p) = \inf\{x: F(x) \ge p\}$ for any distribution function F. Note that $\widehat{F}_k^{-1}(p)$ is simply the p-th quantile of the distribution function $\widehat{F}_k$. T_k is also called the L-estimator of $T(F_k) + \sum_{t=1}^{m} a_t F_k^{-1}(p_t)$. If $a_t = 0$ for all t, then $T_k = T(\widehat{F}_k)$ is called a smooth L-statistic (an L-statistic with smooth weights). If $J(t) \equiv 0$, then $T_k = \sum_{t=1}^{m} a_t \widehat{F}_k^{-1}(p_t)$ is a linear combination of sample quantiles and will be called a non-smooth L-statistic.

Let $\{y_{(l)}, l = 1, \ldots, \sum_{h=1}^{L} \sum_{i=1}^{n_h} n_{hi}\}$ be the order statistics of the sample $\{y_{hij}, j = 1, \ldots, n_{hi}; i = 1, \ldots, n_h; h = 1, \ldots, L\}$ and $\omega_l = w_{hij}/q_k$ if $y_{(l)} = y_{hij}$. Then

$$T(\widehat{F}_k) = \sum_{l} c_l y_{(l)}, \qquad c_l = \omega_l J\left(\sum_{t=1}^{l} \omega_t\right) \tag{2.8}$$

and T_k is a linear combination of the order statistics.

We now discuss some examples of useful L-statistics in survey problems. Obviously the (weighted) sample mean ($J \equiv 1$ and $a_t \equiv 0$) and the sample quantile ($J \equiv 0$) are the most commonly used L-statistics. Some other examples are given as follows.

EXAMPLE 1 (Trimmed sample mean). Let $a_t \equiv 0$ and $J(t) = (\beta - \alpha)^{-1} I(\alpha \le t \le \beta)$, where $0 \le \alpha < \beta \le 1$. The resulting T_k is then a trimmed sample mean. T_k is not necessarily asymptotically normal if $\{F_k\}$ does not have a continuous

and differentiable limit [Stigler (1973)]. Hence we may smooth $J(t)$ and use

$$\tilde{J}(t) = \begin{cases} \dfrac{2(t-\alpha)}{(a-\alpha)(\beta-\alpha+b-a)}, & \alpha \le t < a, \\ \dfrac{2}{(\beta-\alpha+b-a)}, & a \le t < b, \\ \dfrac{2(\beta-t)}{(\beta-b)(\beta-\alpha+b-a)}, & b \le t < \beta, \\ 0, & \text{otherwise}, \end{cases}$$

where a and b are constants satisfying $\alpha < a \le b < \beta$.

EXAMPLE 2 (Weighted decile mean and variance). In the problems where we study income shares for different quantile groups we need to consider the (weighted) mean and variance of a given group. Suppose that we divide the sample into $D = 10$ ordered decile groups. For $d = 1, \ldots, D-1$, let $\hat{\mu}_d = \sum_l c_{dl} y_{(l)}/\lambda_d$ and $\hat{\rho}_d^2 = \sum_l c_{dl}[y_{(l)} - \hat{\mu}_d]^2/\lambda_d$, where $\lambda_d = \sum_l c_{dl}$ and c_{dl} is generated according to (2.8) with the function

$$J(t) = J_d(t) = \begin{cases} 1, & \dfrac{d-1}{D} < t \le \dfrac{d}{D}, \\ 0, & \text{otherwise}, \end{cases} \qquad d = 1, \ldots, D.$$

The decile mean $\hat{\mu}_d$ and variance $\hat{\rho}_d^2$ are functions of λ_d, $\xi_{1d} = \sum_l c_{dl} y_{(l)}$ and $\xi_{2d} = \sum_l c_{dl} y_{(l)}^2$; ξ_{1d} is an L-statistic; λ_d and ξ_{2d} are L-statistics of the more general form considered by Chernoff, Gastwirth and Johns (1967) with $J = J_d$ but with the characteristic $y_{(l)}$ replaced by 1 and $y_{(l)}^2$, respectively.

EXAMPLE 3 (Lorenz curve). The Lorenz curve (or curve of concentration) is one of the most frequently used devices to describe inequality in income or wealth distributions [Beach and Kaliski (1986)]. The sample Lorenz curve ordinates are expressed as $\hat{\gamma}_d/(D\hat{\gamma}_D)$, where $\hat{\gamma}_1 = \hat{\mu}_1, \hat{\gamma}_d = (d-1/d)\hat{\gamma}_{d-1} + (1/d)\hat{\mu}_d$, $d = 2, \ldots, D$, and $\hat{\mu}_d$ are given in Example 2. The $\hat{\gamma}_d$ are functions of smooth L-statistics.

EXAMPLE 4 (Gini's family). The sample Gini's mean difference is a well-known smooth L-statistic with $J(t) = 4t - 2$ [Serfling (1980), Example 8.2.4B]. A related quantity is the sample Gini's coefficient (a measure of variability), which is equal to

$$\frac{[\text{sample Gini's mean difference} - \int x\, d\hat{F}_k(x)/M]}{\int x\, d\hat{F}_k(x)}.$$

This is a function of smooth L-statistics.

More generally, in survey problems we may consider a family of parameters called the Gini's family [Nygard and Sandström (1985)]:

$$GI_k = \frac{\int xJ(F_k(x))\,dF_k(x)}{\int x\,dF_k(x)}. \tag{2.9}$$

For example, when $J(t) = 1 - 3(1-t)^2$, GI_k in (2.9) is the Mehran's measure; when $J(t) = (3t^2 - 1)/2$, GI_k is the Piesch's measure. The sampled version of GI_k is a function of smooth L-statistics.

3. Results for smooth L-statistics. We establish asymptotic normality of smooth L-statistics in this section and will treat nonsmooth L-statistics in the next section. The results can then be extended to general L-statistics of the form (2.7) in a straightforward manner. For smooth L-statistics $T(\widehat{F}_k)$ with T given by (2.6), we can write [see, e.g., Serfling (1980), Chapter 8]

$$T(\widehat{F}_k) - T(F_k) = \int \phi(x, F_k)\,d\widehat{F}_k(x) - \int Q_k(x)\big[\widehat{F}_k(x) - F_k(x)\big]\,dx,$$

where

$$Q_k(x) = \begin{cases} \dfrac{\psi(\widehat{F}_k(x)) - \psi(F_k(x))}{\widehat{F}_k(x) - F_k(x)} - J(F_k(x)), & \text{if } \widehat{F}_k(x) \neq F_k(x), \\ 0, & \text{if } \widehat{F}_k(x) = F_k(x), \end{cases}$$

$\psi(t) = \int_0^t J(s)\,ds$, and

$$\phi(x, F_k) = -\int \big[I(x \le y) - F_k(y)\big]J(F_k(y))\,dy$$

is the influence function. Let $z_{hij} = \phi(y_{hij}, F_k)$ and $Z_{hij} = \phi(Y_{hij}, F_k)$. Then

$$\int \phi(x, F_k)\,d\widehat{F}_k(x) = \sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{n_{hi}} \frac{w_{hij}z_{hij}}{q_k} = \frac{\bar{z}}{q_k} \tag{3.1}$$

and

$$E\bar{z} = \frac{1}{M}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{j=1}^{N_{hi}} Z_{hij} = 0.$$

There are two types of J-functions, trimmed and untrimmed J, and we will treat them separately.

3.1. *Trimmed L-statistics.* A trimmed J satisfies

$$J(t) = 0 \quad \text{if } t < \alpha \text{ or } t > \beta, \tag{3.2}$$

where α and β are constants satisfying $0 < \alpha < \beta < 1$. The corresponding L-statistic is called a trimmed L-statistic. The trimmed sample mean in Example 1 and the weighted decile mean and variance in Example 2 are trimmed L-statistics. Let

$$\sigma_k^2 = \mathrm{Var}\left(\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{n_{hi}} w_{hij} z_{hij}\right). \tag{3.3}$$

THEOREM 1. *Suppose that assumption* (A1) *and the following assumptions hold*:

(A2) $\liminf_{k\to\infty} n\sigma_k^2 > 0$.

(A3) *The function J is trimmed* [*i.e.*, *J satisfies* (3.2)], *bounded and $m(D) = 0$, where m is the Lebesque measure and $D = \{x: J$ is discontinuous at a limit point of $\{F_k(x), k = 1, 2, \ldots\}\}$.*

(A4) *There are constants c_α and c_β such that*

$$\sup_k F_k(c_\alpha) < \alpha \quad \text{and} \quad \inf_k F_k(c_\beta) > \beta.$$

Then

$$T(\widehat{F}_k) = T(F_k) + \bar{z}/q_k + o_p(n^{-1/2}) \tag{3.4}$$

and

$$\left[T(\widehat{F}_k) - T(F_k)\right]/\sigma_k \to_{\mathcal{L}} N(0,1).$$

REMARKS.

(i) As we discussed in Section 2, the condition $m(D) = 0$ in (A3) cannot be relaxed. For the trimmed mean with $J(t) = (\beta - \alpha)^{-1} I(\alpha \le t \le \beta), D = \{x: \alpha$ or β is a limit point of $\{F_k(x)\}\}$ and J satisfies (A3) if $m(D) = 0$. Assumption (A3) is not needed if we use the smoothly trimmed sample mean in Example 1.

(ii) Assumption (A4) is much weaker than the tightness of the sequence $\{F_k\}$.

PROOF OF THEOREM 1. Under assumptions (A3) and (A4), $\{z_{hij}\}$ is bounded. Under assumptions (A1)–(A4), Liapounov's condition holds for the "weighted average" $\bar{z}$ in (3.1). Thus, by Bickel and Freedman [(1984), Theorem 3] or Krewski and Rao [(1981), Theorem 3.1] $(\bar{z} - E\bar{z})/\sigma_k \to_{\mathcal{L}} N(0, 1)$. Since $E\bar{z} = 0$ and $q_k \to_p 1$, $z/(q_k\sigma_k) \to_{\mathcal{L}} N(0, 1)$ and the second assertion of the theorem follows from the first assertion. Since

$$T(\widehat{F}_k) - T(F_k) = \bar{z}/q_k - \int Q_k(x)\left[\widehat{F}_k(x) - F_k(x)\right] dx,$$

the result follows if

$$n^{1/2} \int Q_k(x)\left[\widehat{F}_k(x) - F_k(x)\right] dx \to_p 0. \tag{3.5}$$

Define

$$\mathcal{A}_k = \{\widehat{F}_k(c_\beta) > \beta \text{ and } \widehat{F}_k(c_\alpha) < \alpha\}.$$

From assumption (A4) and the fact that $\widehat{F}_k(c_\alpha) - F_k(c_\alpha) \to_p 0$ and $\widehat{F}_k(c_\beta) - F_k(c_\beta) \to_p 0$, we have $P(\mathcal{A}_k) \to 1$. On the event $\mathcal{A}_k$,

$$\begin{aligned} &\left| \int Q_k(x)[\widehat{F}_k(x) - F_k(x)]\, dx \right| \\ &\quad = \left| \int_{c_\alpha}^{c_\beta} Q_k(x)[\widehat{F}_k(x) - F_k(x)]\, dx \right| \\ &\quad \le \left\{ \int_{c_\alpha}^{c_\beta} [Q_k(x)]^2\, dx \int_{c_\alpha}^{c_\beta} [\widehat{F}_k(x) - F_k(x)]^2\, dx \right\}^{1/2}. \end{aligned} \tag{3.6}$$

From the assumption (A3) and $\widehat{F}_k(x) - F_k(x) \to_p 0$, $Q_k(x) \to_p 0$ for $x \notin D$. Also, $\|Q_k\|_\infty \le 2\|J\|_\infty$. Hence

$$\int_{c_\alpha}^{c_\beta} [Q_k(x)]^2\, dx \to_p 0. \tag{3.7}$$

For any x, $E[G_k(x) - F_k(x)]^2 = \mathrm{Var}[G_k(x)] \le \max_{j,i,h} n_{hi} w_{hij} \le cn^{-1}$ by assumption (A1), where c is a constant. Hence

$$E \int_{c_\alpha}^{c_\beta} [G_k(x) - F_k(x)]^2\, dx \le \frac{c(c_\beta - c_\alpha)}{n} = O(n^{-1}). \tag{3.8}$$

Since $\widehat{F}_k = G_k/q_k$, we have

$$\begin{aligned} \int_{c_\alpha}^{c_\beta} [\widehat{F}_k(x) - F_k(x)]^2\, dx &\le 2 \int_{c_\alpha}^{c_\beta} [G_k(x) - F_k(x)]^2\, dx + 2 \int_{c_\alpha}^{c_\beta} [\widehat{F}_k(x) - G_k(x)]^2\, dx \\ &\le 2 \int_{c_\alpha}^{c_\beta} [G_k(x) - F_k(x)]^2\, dx + 2(q_k - 1)^2 (c_\beta - c_\alpha) \\ &= O_p(n^{-1}), \end{aligned}$$

by (3.8) and $q_k - 1 = O_p(n^{-1/2})$. This, together with (3.6) and (3.7), implies (3.5). The proof is complete. □

3.2. *Untrimmed L-statistics.* For untrimmed L-statistics, some moment conditions are necessary. We only consider smooth functions J.

THEOREM 2. *Suppose that assumptions* (A1), (A2) *and the following hold:*

(A3′) *The function J is Lipschitz continuous on* $[0, 1]$;
(A4′) $\sup_k \int_{-\infty}^{\infty} |x|\, dF_k(x) < \infty$ *and there is a $\delta > 0$ such that*

$$n^{1+\delta} \sum_{h=1}^{L} \sum_{i=1}^{n_h} E|u_{hi} - Eu_{hi}|^{2(1+\delta)} \to 0, \tag{3.9}$$

where $u_{hi} = \sum_{j=1}^{n_{hi}} w_{hij} z_{hij}$.

Then the conclusions of Theorem 1 *still hold.*

REMARK. Note that (3.9) is a Liapounov-type condition similar to condition C1 in Krewski and Rao (1981).

PROOF OF THEOREM 2. Under (A2) and (3.9), $(\bar{z} - E\bar{z})/\sigma_k \to_{\mathcal{L}} N(0,1)$. Thus, from the proof of Theorem 1, we only need to show that (3.5) holds. Since J is Lipschitz continuous, there is a constant $c_J > 0$ such that $|J(t) - J(s)| \le c_J|t-s|, t, s \in [0,1]$. Then

$$\left| \int Q_k(x)\left[\widehat{F}_k(x) - F_k(x)\right] dx \right| \le c_J \int \left[\widehat{F}_k(x) - F_k(x)\right]^2 dx.$$

From $\widehat{F}_k = G_k/q_k$,

$$\begin{aligned}
&\int_{-\infty}^{0} \left[\widehat{F}_k(x) - F_k(x)\right]^2 dx \\
&\quad \le 2\int_{-\infty}^{0} \left[G_k(x) - F_k(x)\right]^2 dx + 2\int_{-\infty}^{0} \left[\widehat{F}_k(x) - G_k(x)\right]^2 dx \\
&\quad = 2\int_{-\infty}^{0} \left[G_k(x) - F_k(x)\right]^2 dx + 2\left(q_k^{-1} - 1\right)^2 \int_{-\infty}^{0} \left[G_k(x)\right]^2 dx.
\end{aligned}$$

Note that $(q_k^{-1} - 1)^2 = q_k^{-2}(q_k - 1)^2 = O_p(n^{-1})$,

$$E\int_{-\infty}^{0} \left[G_k(x)\right]^2 dx \le \int_{-\infty}^{0} E\left[G_k(x)\right] dx = \int_{-\infty}^{0} F_k(x)\, dx \le \int |x|\, dF_k(x)$$

and

$$\begin{aligned}
E\int_{-\infty}^{0} \left[G_k(x) - F_k(x)\right]^2 dx &= \int_{-\infty}^{0} \operatorname{Var}[G_k(x)]\, dx \\
&\le \int_{-\infty}^{0} \sum_{h=1}^{L} \sum_{i=1}^{n_h} E\left[\sum_{j=1}^{n_{hi}} w_{hij} I(y_{hij} \le x)\right]^2 dx \\
&\le O(n^{-1}) \int_{-\infty}^{0} F_k(x)\, dx = O(n^{-1}),
\end{aligned}$$

where the last inequality follows from assumption (A1). Hence

$$\int_{-\infty}^{0} \left[\widehat{F}_k(x) - \widehat{F}_k(x)\right]^2 dx = O_p(n^{-1}).$$

Similarly,

$$\int_{0}^{\infty} \left[\widehat{F}_k(x) - \widehat{F}_k(x)\right]^2 dx = O_p(n^{-1}),$$

and therefore (3.5) follows. This completes the proof. □

The J-functions corresponding to the untrimmed L-statistics in Examples 1–4 are clearly Lipschitz continuous on $[0, 1]$.

In some cases we need to consider a statistic S_k which is a function of several L-statistics (Examples 2–4), say $T_k^{(j)}, j = 1, \ldots, m$. The asymptotic distribution of S_k can be obtained by using the results in Theorems 1 and 2. Since each $T_k^{(j)}$ can be approximated by a weighted average according to (3.4), we can apply the central limit theorem and the δ-method. The details are omitted.

3.3. *Variance estimation by jackknife.* For various purposes in statistical analysis we need to estimate the unknown asymptotic variance of a given statistic. When the given statistic is a smooth L-statistic, a formula for its asymptotic variance is given by (3.3). When the given statistic is a known function of several L-statistics, a formula for its asymptotic variance can be obtained by using the δ-method and calculating partial derivatives of the known function. In some cases, the asymptotic variance can be estimated by substituting the unknown quantities in the formula of the asymptotic variance by some estimators. This is called the substitution method. Another very popular method in sample surveys is the jackknife [Tukey (1958)]. Unlike the substitution method, the jackknife does not require knowing a formula of the asymptotic variance, which is one of the reasons why the jackknife is so attractive. Note that for statistics such as the sample Lorenz curve ordinates (Example 3), the derivation of their asymptotic variance is quite complicated and tedious. The jackknife replaces the theoretical derivation by repeated computations of the given statistic. Many agencies (e.g., Statistics Canada) have computer software to implement the computation of the jackknife variance estimators.

Variance estimation by jackknife is based on the variability among a number of replicate statistics computed from overlapping subsamples of the total sample. Let

$$G_k^{(gl)}(x) = \sum_{h \neq g} \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} w_{hij} I(y_{hij} \leq x) + \frac{n_g}{n_g - 1} \sum_{i \neq l} \sum_{j=1}^{n_{gi}} w_{gij} I(y_{gij} \leq x),$$

$$q_k^{(gl)} = G_k^{(gl)}(\infty), \qquad \widehat{F}_k^{(gl)} = \frac{\widehat{G}_k^{(gl)}}{q_k^{(gl)}}$$

$$T_k^{(gl)} = T\big(\widehat{F}_k^{(gl)}\big). \qquad l = 1; \ldots, n_g; g = 1, \ldots, L.$$

The jackknife estimator of the asymptotic variance of $T_k = T(\widehat{F}_k)$ is

$$\widehat{\sigma}_k^2 = \sum_{h=1}^{L} \frac{(1 - f_h)(n_h - 1)}{n_h} \sum_{i=1}^{n_h} \big(T_k^{(hi)} - T_k\big)^2, \tag{3.10}$$

where $f_h = n_h / N_h$ if the first-stage sampling is without replacement and $f_h = 0$ if the first-stage sampling is with replacement. When the statistic T_k is a function

of several L-statistics, its jackknife estimator of the asymptotic variance is still given by (3.10) with $T_k^{(hi)}$ being defined as the statistic based on the subsample obtained by removing the ith cluster of the hth stratum.

We now establish the consistency of the jackknife variance estimator $\hat{\sigma}_k^2$ in (3.10). First, consider the case where $T_k = T(\hat{F}_k)$.

THEOREM 3 (Trimmed L-statistics). *Suppose that assumptions* (A1)–(A4) *and the following hold*:

(A5) *There are two sequence of sets* $\{\mathcal{N}_{1k}\}$ *and* $\{\mathcal{N}_{2k}\}$ *such that, for each* k,

$$\{n_h, h = 1, \ldots, L\} = \mathcal{N}_{1k} \cup \mathcal{N}_{2k},$$

$$\sup_k \max\{n_h \in \mathcal{N}_{1k}\} < \infty \quad \text{and} \quad \lim_{k \to \infty} \min\{n_h \in \mathcal{N}_{2k}\} = \infty.$$

Then the jackknife estimator $\hat{\sigma}_k^2$ *in* (3.10) *is consistent for* σ_k^2 *in* (3.3), *that is*,

$$\hat{\sigma}_k^2 / \sigma_k^2 \to_p 1.$$

REMARK. Assumption (A5) is satisfied in the following two common situations in surveys: (i) all the n_h are small (bounded by a constant); (ii) all the n_h are large. Assumption (A5) allows a mixture of some small n_h's and some large n_h's.

PROOF OF THEOREM 3. Let a_{hij} denote the function $a_{hij}(x) = w_{hij} I(y_{hij} \le x)$. Then

$$\begin{aligned}\hat{F}_k^{(hi)} - \hat{F}_k &= \frac{G_k^{(hi)}}{q_k^{(hi)}} - \frac{G_k}{q_k} = G_k^{(hi)}\left(\frac{1}{q_k^{(hi)}} - \frac{1}{q_k}\right) + \frac{1}{q_k}\left(G_k^{(hi)} - G_k\right) \\ &= G_k^{(hi)}\left(\frac{1}{q_k^{(hi)}} - \frac{1}{q_k}\right) + \frac{n_h}{q_k(n_h - 1)}\left(\frac{1}{n_h}\sum_{i=1}^{n_h}\sum_{j=1}^{n_{hi}} a_{hij} - \sum_{j=1}^{n_{hi}} a_{hij}\right).\end{aligned}$$

Let u_{hi} be defined in (3.9), $\bar{u}_h = n_h^{-1}\sum_{i=1}^{n_h} u_{hi}$ and $\bar{z}^{(hi)} = \int \phi(x, F_k)\, dG_k^{(hi)}(x)$. Then

$$\begin{aligned}T_k^{(hi)} - T_k = {} & \frac{n_h}{q_k(n_h - 1)}(\bar{u}_h - u_{hi}) + \bar{z}^{(hi)}\left(\frac{1}{q_k^{(hi)}} - \frac{1}{q_k}\right) \\ & - \int Q_k^{(hi)}(x)\left[\hat{F}_k^{(hi)}(x) - \hat{F}_k(x)\right] dx,\end{aligned}$$

where

$$Q_k^{(hi)}(x) = \begin{cases} \dfrac{\psi\left(\hat{F}_k^{(hi)}(x)\right) - \psi\left(\hat{F}_k(x)\right)}{\hat{F}_k^{(hi)}(x) - \hat{F}_k(x)} - J\left(F_k(x)\right), & \text{if } \hat{F}_k^{(hi)}(x) \ne \hat{F}_k(x), \\ 0, & \text{otherwise,} \end{cases}$$

and $\psi(t) = \int_0^t J(s)\,ds$. It follows from (3.10) that

$$\widehat{\sigma}_k^2 = \alpha_{1k} + \alpha_{2k} + \alpha_{3k} + 2(\text{cross-product terms}), \tag{3.11}$$

where

$$\alpha_{1k} = \sum_{h=1}^{L} \frac{(1-f_h)n_h}{n_h - 1} \sum_{i=1}^{n_h} \frac{(u_{hi} - \overline{u}_h)^2}{q_k^2}, \tag{3.12}$$

$$\alpha_{2k} = \sum_{h=1}^{L} \frac{(1-f_h)(n_h-1)}{n_h} \sum_{i=1}^{n_h} \left(\overline{z}_h^{(hi)}\right)^2 \left(\frac{1}{q_k^{(hi)}} - \frac{1}{q_k}\right)^2, \tag{3.13}$$

$$\alpha_{3k} = \sum_{h=1}^{L} \frac{(1-f_h)(n_h-1)}{n_h} \sum_{i=1}^{n_h} \left(r_k^{(hi)}\right)^2 \tag{3.14}$$

and

$$r_k^{(hi)} = -\int Q_k^{(hi)}(x)\left[\widehat{F}_k^{(hi)}(x) - \widehat{F}_k(x)\right] dx. \tag{3.15}$$

To establish the result, it suffices to show that

$$n\left(\alpha_{1k} - \sigma_k^2\right) \to_p 0, \tag{3.16}$$

$$n\alpha_{2k} \to_p 0 \quad \text{and} \quad n\alpha_{3k} \to_p 0, \tag{3.17}$$

since under (3.16) and (3.17), the cross-product terms are of the order $o_p(n^{-1})$.

PROOF OF (3.16). Let

$$\sigma_{1k}^2 = \sum_{n_h \in \mathcal{N}_{1k}} \operatorname{Var}\left(\sum_{i=1}^{n_h} u_{hi}\right) \quad \text{and} \quad \sigma_{2k}^2 = \sum_{n_h \in \mathcal{N}_{2k}} \operatorname{Var}\left(\sum_{i=1}^{n_h} u_{hi}\right).$$

Then $\sigma_k^2 = \sigma_{1k}^2 + \sigma_{2k}^2$ and

$$E\left[\sum_{n_h \in \mathcal{N}_{1k}} \frac{n(1-f_h)n_h}{n_h - 1} \sum_{i=1}^{n_h} (u_{hi} - \overline{u}_h)^2\right] = n\sigma_{1k}^2.$$

Under assumptions (A1) and (A5), there is a constant $c > 0$ such that

$$n^{1+\delta} \sum_{n_h \in \mathcal{N}_{1k}} E\left|\frac{(1-f_h)n_h}{n_h - 1} \sum_{i=1}^{n_h} (u_{hi} - \overline{u}_h)^2\right|^{1+\delta} \le cn^{1+\delta} \sum_{n_h \in \mathcal{N}_{1k}} \sum_{i=1}^{n_h} E u_{hi}^{2(1+\delta)} \to 0.$$

Hence

$$\sum_{n_h \in \mathcal{N}_{1k}} \frac{n(1-f_h)n_h}{n_h - 1} \sum_{i=1}^{n_h} (u_{hi} - \overline{u}_h)^2 - n\sigma_{1k}^2 \to_p 0. \tag{3.18}$$

Let $\mu_h = Eu_{hi}$. Since

$$E\left[\sum_{n_h \in \mathcal{N}_{2k}} \frac{n(1-f_h)n_h^2}{n_h-1}(\mu_h - \overline{u}_h)^2\right]$$
$$= \sum_{n_h \in \mathcal{N}_{2k}} \frac{n(1-f_h)n_h^2}{n_h-1} \operatorname{Var}(\overline{u}_h)$$
$$\le \frac{n}{\min\{n_h \in \mathcal{N}_{2k}\}-1} \sum_{n_h \in \mathcal{N}_{2k}} \operatorname{Var}\left(\sum_{i=1}^{n_h} u_{hi}\right) \to 0$$

[by assumptions (A1)–(A5)] and

$$E\left[\sum_{n_h \in \mathcal{N}_{2k}} \frac{n(1-f_h)}{n_h-1}\sum_{i=1}^{n_h}(u_{hi} - \mu_h)^2\right]$$
$$\le \frac{n}{\min\{n_h \in \mathcal{N}_{2k}\}-1} \sum_{n_h \in \mathcal{N}_{2k}} \operatorname{Var}\left(\sum_{i=1}^{n_h} u_{hi}\right) \to 0,$$

we have

$$\sum_{n_h \in \mathcal{N}_{2k}} \frac{n(1-f_h)n_h}{n_h-1}\sum_{i=1}^{n_h}(u_{hi} - \overline{u}_h)^2 = \sum_{n_h \in \mathcal{N}_{2k}} \frac{n(1-f_h)n_h}{n_h-1}\sum_{i=1}^{n_h}(u_{hi} - \mu_h)^2$$
$$- \sum_{n_h \in \mathcal{N}_{2k}} \frac{n(1-f_h)n_h^2}{n_h-1}\sum_{i=1}^{n_h}(\mu_h - \overline{u}_h)^2$$
$$= \sum_{n_h \in \mathcal{N}_{2k}} n(1-f_h)\sum_{i=1}^{n_h}(u_{hi} - \mu_h)^2 + o_p(1).$$

Since

$$E\left[\sum_{n_h \in \mathcal{N}_{2k}} n(1-f_h)\sum_{i=1}^{n_h}(u_{hi} - \mu_h)^2\right] = n\sigma_{2k}^2 + o(1)$$

and

$$n^{1+\delta} \sum_{n_h \in \mathcal{N}_{2k}} \sum_{i=1}^{n_h}(1-f_h)^{1+\delta} E|u_{hi} - \mu_h|^{2(1+\delta)} \to 0,$$

we have

$$\sum_{n_h \in \mathcal{N}_{2k}} \frac{n(1-f_h)n_h}{n_h-1}\sum_{i=1}^{n_h}(u_{hi} - \overline{u}_h)^2 - n\sigma_{2k}^2 \to 0. \tag{3.19}$$

Since $q_k^2 \to_p 1$, (3.16) follows from (3.12), (3.18) and (3.19). □

PROOF OF (3.17). Let

$$\zeta_k = \sum_{h=1}^{L} \frac{(1-f_h)(n_h-1)}{n_h} \sum_{i=1}^{n_h} \left(\frac{1}{q_k^{(hi)}} - \frac{1}{q_k} \right)^2.$$

Then ζ_k is the jackknife estimator of the asymptotic variance of q_k^{-1}. Thus $n\zeta_k = O_p(1)$. Since $\bar{z}^{(hi)}$ are weighted averages, it can be shown that $\max_{h,i} |\bar{z}^{(hi)} - \bar{z}| \to_p 0$; but $E\bar{z} = 0$ and $\bar{z} - E\bar{z} \to_p 0$. Hence, from (3.13),

$$n\alpha_{2k} \le \max_{h,i} \left(\bar{z}^{(hi)}\right)^2 n\zeta_k = o_p(1).$$

Note that $\max_{h,i} \|\widehat{F}_k^{(hi)} - \widehat{F}_k\|_\infty \to 0$. Hence, similar to the proof of Theorem 1,

$$\left(r_k^{(hi)}\right)^2 \le \int_{c_\alpha}^{c_\beta} \left[Q_k^{(hi)}(x)\right]^2 dx \int_{c_\alpha}^{c_\beta} \left[\widehat{F}_k^{(hi)}(x) - \widehat{F}_k(x)\right]^2 dx$$

and

$$\max_{h,i} \int_{c_\alpha}^{c_\beta} \left[Q_k^{(hi)}(x)\right]^2 dx \to_p 0.$$

From (3.14) and (3.15),

$$\begin{aligned} n\alpha_{3k} &\le \left(\max_{h,i} \int_{c_\alpha}^{c_\beta} \left[Q_k^{(hi)}(x)\right]^2 dx \right) n \sum_{h=1}^{L} \frac{(1-f_h)(n_h-1)}{n_h} \\ &\quad \times \sum_{i=1}^{n_h} \int_{c_\alpha}^{c_\beta} \left[\widehat{F}_k^{(hi)}(x) - \widehat{F}_k(x)\right]^2 dx \\ &\le o_p(1) \left[n\zeta_k + n \sum_{h=1}^{L} \frac{(1-f_h)n_h}{n_h - 1} \frac{1}{q_k^2} \sum_{i=1}^{n_h} \left(\sum_{j=1}^{n_{hi}} w_{hij} \right)^2 \right] = o_p(1), \end{aligned}$$

since, by (A1),

$$n \sum_{h=1}^{L} \sum_{i=1}^{n_h} \left(\sum_{j=1}^{n_{hi}} w_{hij} \right)^2 \le n \max_{h,i,j} (n_{hi} w_{hij}) \sum_{h=1}^{L} \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} w_{hij} = O_p(1).$$

This proves (3.17). The proof of Theorem 3 now is complete. □

THEOREM 4 (Untrimmed L-statistics). *The conclusion of Theorem* 3 *holds also for untrimmed L-statistics if assumptions* (A3) *and* (A4) *are replaced by* (A3′) *and* (A4′), *respectively.*

PROOF. Note that (3.11)–(3.15) still hold and we need to show (3.16) and (3.17). Using (A4′) and a proof similar to that of Theorem 3, we can show (3.16)

and the first assertion of (3.17). It remains to show the second assertion of (3.17). From the Lipschitz continuity of J,

$$(r_k^{(hi)})^2 \le c_J^2 \left\{ \int [\widehat{F}_k^{(hi)}(x) - \widehat{F}_k(x)]^2 \, dx \right\}^2.$$

Note that $\max_{h,i} |1/q_k^{(hi)} - 1/q_k| = o_p(1)$ and $\max_{h,i} 1/q_k^{(hi)} = O_p(1)$. Then

$$\begin{aligned}
&\sum_{h=1}^{L} \frac{n(1-f_h)(n_h-1)}{n_h} \sum_{i=1}^{n_h} \left\{ \int_{-\infty}^{0} [G_k(x)]^2 \left(\frac{1}{q_k^{(hi)}} - \frac{1}{q_k} \right)^2 dx \right\}^2 \\
&\quad \le \max_{h,i} \left(\frac{1}{q_k^{(hi)}} - \frac{1}{q_k} \right)^2 \left\{ \int_{-\infty}^{0} [G_k(x)]^2 \, dx \right\}^2 n\zeta_k \\
&\quad = o_p(1) O_p(1) O_p(1) = o_p(1),
\end{aligned}$$

and by (A1) and (A4′),

$$\begin{aligned}
&\sum_{h=1}^{L} \frac{n(1-f_h)(n_h-1)}{n_h} \sum_{i=1}^{n_h} \left(\frac{1}{q_k^{(hi)}} \right)^2 \left\{ \int_{-\infty}^{0} [G_k^{(hi)}(x) - G_k(x)]^2 \, dx \right\}^2 \\
&\quad \le O_p(n^{-1}) \sum_{h=1}^{L} \sum_{i=1}^{n_h} \left[\int_{-\infty}^{0} \sum_{j=1}^{n_{hi}} a_{hij}(x) \, dx \right]^2 \\
&\quad \le O_p(n^{-1}) \left[\int_{-\infty}^{0} G_k(x) \, dx \right]^2 = O_p(n^{-1}).
\end{aligned}$$

From $\widehat{F}_k^{(hi)} - \widehat{F}_k = (G_k^{(hi)} - G_k)/q_k^{(hi)} + G_k(1/q_k^{(hi)} - 1/q_k)$, we have

$$\sum_{h=1}^{L} \frac{n(1-f_h)(n_h-1)}{n_h} \sum_{i=1}^{n_h} \left\{ \int_{-\infty}^{0} [\widehat{F}_k^{(hi)}(x) - \widehat{F}_k(x)]^2 \, dx \right\}^2 \to_p 0.$$

Similarly,

$$\sum_{h=1}^{L} \frac{n(1-f_h)(n_h-1)}{n_h} \sum_{i=1}^{n_h} \left\{ \int_{0}^{\infty} [\widehat{F}_k^{(hi)}(x) - \widehat{F}_k(x)]^2 \, dx \right\}^2 \to_p 0.$$

This proves (3.17) and completes the proof. □

When the statistic of interest is a function of several smooth L-statistics, the consistency of its jackknife variance estimator can be easily established by using the results in Theorems 3 and 4 and the argument used in the proof of Theorem 3.4 of Krewski and Rao (1981), who proved the consistency of the jackknife variance estimator for a function of several sample means. Therefore, our results can be applied to Examples 1–4 for variance estimation.

4. Results for sample quantiles. For nonsmooth L-statistics, we focus on the sample pth quantile with a given $p \in (0, 1)$. The case of functions of several sample quantiles can be treated by using the δ-method.

Let $\theta_k = F_k^{-1}(p)$ be the population pth quantile, and let $\hat{\theta}_k = \hat{F}^{-1}(p)$ be the sample pth quantile. Francisco and Fuller (1991) established a Bahadur-type representation for $\hat{\theta}_k$ in complex survey problems. A similar result was obtained in Shao and Wu (1992) for the case of one-stage stratified simple random smapling. Francisco and Fuller (1991) also proposed an estimator [see (4.15)] of the asymptotoic variance of $\hat{\theta}_k$ and proved its consistency under their conditions 1–7. However, some of these conditions (conditions 5–7) are unnecessary. In this section we derive an estimator of the asymptotic variance of $\hat{\theta}_k$ and establish its consistency under weaker conditions.

Suppose that assumption (A1) and the following assumptions hold.

(A6) The sequence $\{\theta_k\}$ is bounded.

(A7) There is a sequence of functions $\{f_k(\cdot)\}$ such that

$$\lim_{k\to\infty}\left[\frac{F_k(\theta_k+\delta_k)-F_k(\theta_k)}{\delta_k}-f_k(\theta_k)\right]=0, \tag{4.1}$$

for any sequence $\{\delta_k\}$ of the order $O(n^{-1/2})$ and

$$0<\inf_k f_k(\theta_k)\le\sup_k f_k(\theta_k)<\infty.$$

Then, by using a proof similar to those in Francisco and Fuller (1991) and Shao and Wu (1992), we can show that

$$\hat{\theta}_k=\theta_k+\frac{F_k(\theta_k)-\hat{F}_k(\theta_k)}{f_k(\theta_k)}+o_p(n^{-1/2}). \tag{4.2}$$

Furthermore, it can be shown that

$$n^{1/2}\left[\hat{F}_k^{-1}(p_k)-\hat{F}_k^{-1}(p)\right]=\frac{c}{f_k(\theta_k)}+o_p(1), \tag{4.3}$$

where $p_k = p + cn^{-1/2}$ and c is a constant, under the following assumption:

$$\text{(A8)}\quad \lim_{k\to\infty}\left[\frac{F_k^{-1}(p_k)-F_k^{-1}(p)}{p_k-p}-\frac{1}{f_k(\theta_k)}\right]=0 \text{ and } \lim_{k\to\infty}\left|f_k\left(F_k^{-1}(p)\right)-f_k(\theta_k)\right|=0.$$

Assumptions (A7) and (A8) are essentially the same as Condition 4 in Francisco and Fuller (1991). Condition (4.1) is a type of smoothness condition requiring that F_k is nearly differentiable at θ_k when k is large, although F_k is not differentiable for each fixed k. When the units in the finite populations $\mathcal{P}_k$ are samples from continuous and differentiable superpopulations, (4.1) is satisfied.

As a direct consequence of (4.2),

$$\frac{\hat{\theta}_k - \theta_k}{\sigma_k} \to_{\mathcal{L}} N(0,1), \tag{4.4}$$

where

$$\sigma_k^2 = \frac{v_k(\theta_k)}{[f_k(\theta_k)]^2} \tag{4.5}$$

is the asymptotic variance of $\hat{\theta}_k$,

$$v_k(\theta_k) = F_k^2(\theta_k)\,\mathrm{Var}(q_k) - 2F_k(\theta_k)\,\mathrm{Cov}[q_k, G_k(\theta_k)] + \mathrm{Var}\,[G_k(\theta_k)] \tag{4.6}$$

is the asymptotic variance of $\hat{F}_k(\theta_k)$, and $\liminf_k nv_k(\theta_k) > 0$ is assumed.

To estimate σ_k^2 in (4.5), however, is much more difficult (or requires much stronger conditions) than to estimate the asymptotic variance of a smooth L-statistic. Unlike the case of smooth L-statistics, the jackknife variance estimator may not work for sample quantiles. More complicated data-resampling methods such as the delete-d jackknife [Shao and Wu (1989)], the balanced repeated replication [McCarthy (1969) and Shao and Wu (1992)] and the bootstrap [Efron (1979)] need to be considered. The estimator of σ_k^2 considered here is of the form

$$\hat{\sigma}_k^2 = \hat{v}_k(\hat{\theta}_k)\hat{u}_k^2, \tag{4.7}$$

where

$$\hat{u}_k = n^{1/2}[\hat{F}_k^{-1}(p + n^{-1/2}) - \hat{F}_k^{-1}(p - n^{-1/2})]/2 \tag{4.8}$$

is a consistent estimator of $1/f_k(\theta_k)$ by (4.3),

$$\hat{v}_k(x) = \frac{\hat{F}_k^2(x)\widehat{\mathrm{var}}(\infty) - 2\hat{F}_k(x)\widehat{\mathrm{cov}}(x) + \widehat{\mathrm{var}}(x)}{q_k^2},$$

$$\widehat{\mathrm{var}}(x) = \sum_{h=1}^{L} \frac{(1-f_h)n_h}{n_h - 1} \sum_{l=1}^{n_h} \left[a_{hi}(x) - \frac{1}{n_h} \sum_{l=1}^{n_h} a_{hl}(x) \right]^2,$$

$$\widehat{\mathrm{cov}}(x) = \sum_{h=1}^{L} \frac{(1-f_h)n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[a_{hi}(x) - \frac{1}{n_h} \sum_{i=1}^{n_h} a_{hl}(x) \right] \left[a_{hi}(\infty) - \frac{1}{n_h} \sum_{l=1}^{n_h} a_{hl}(\infty) \right]$$

and $a_{hi}(x) = \sum_{j=1}^{n_{hi}} a_{hij}(x) = \sum_{j=1}^{n_{hi}} w_{hij} I(y_{hij} \le x)$.

Note that, for each fixed x, $\widehat{\mathrm{var}}(x)$ and $\widehat{\mathrm{cov}}(x)$ are unbiased and consistent estimators of $\mathrm{Var}[G_k(x)]$ and $\mathrm{Cov}[q_k, G_k(x)]$, respectively. Naturally, it is expected that $\hat{v}_k(\theta_k)$ is consistent for $v_k(\theta_k)$ in (4.6); but $\hat{v}_k(\theta_k)$ is not an estimator since θ_k is unknown. We have to replace θ_k by $\hat{\theta}_k$ and use $\hat{v}_k = \hat{v}_k(\hat{\theta}_k)$ as an estimator of $v_k(\theta_k)$. The estimator $\hat{v}_k$ was also used by Francisco and Fuller (1991). The following result shows the consistency of $\hat{v}_k$ and $\hat{\sigma}_k^2$.

THEOREM 5. *Assume* (A1), (A6) *and the following*:

(A9) *For any* $x \in \Theta$,

$$\lim_{t \to 0} \lim_{k \to \infty} \left[F_k(x+t) - F_k(x)\right] = 0,$$

where Θ *is the set of limit points of the sequence* $\{\theta_k\}$. *Then*

$$n\left[\hat{v}_k - v_k(\theta_k)\right] \to_p 0. \tag{4.9}$$

If (A7) *and* (A8) *also hold, then* $\hat{\sigma}_k^2$ *in* (4.7) *is consistent for* σ_k^2 *in* (4.5), *that is*,

$$\hat{\sigma}_k^2 / \sigma_k^2 \to_p 1. \tag{4.10}$$

REMARK. Similar to (4.1), assumption (A9) requires that F_k is nearly continuous at each $x \in \Theta$ when k is large; (A9) holds if $\{F_k\}$ has a continuous limit. If $\theta_k = \theta$ for all k, then (A9) is implied by (4.1).

PROOF OF THEOREM 5. Since (4.10) follows from (4.3), (4.7) and (4.9), we only need to show (4.9). Using a subsequence argument, we can assume that there is a $\theta \in \Theta$ such that

$$\theta_k \to \theta \quad \text{and} \quad \hat{\theta}_k \to \theta \quad \text{a.s.}$$

Define

$$\hat{\tau}_{1k}(x) = \sum_{h=1}^{L} \frac{(1-f_h)n_h}{n_h - 1} \sum_{i=1}^{n_h} a_{hi}^2(x),$$

$$\hat{\tau}_{2k}(x) = \sum_{h=1}^{L} \frac{1-f_h}{n_h - 1} \left[\sum_{i=1}^{n_h} a_{hi}(x)\right]^2,$$

$$\hat{\tau}_{3k}(x) = \sum_{h=1}^{L} \frac{(1-f_h)n_h}{n_h - 1} \sum_{i=1}^{n_h} a_{hi}(x) a_{hi}(\infty),$$

$$\hat{\tau}_{4k}(x) = \sum_{h=1}^{L} \frac{1-f_h}{n_h - 1} \left[\sum_{i=1}^{n_h} a_{hi}(x)\right] \left[\sum_{i=1}^{n_h} a_{hi}(\infty)\right]$$

and

$$\tau_{1k}(x) = E\hat{\tau}_{lk}(x), \qquad l = 1, \ldots, 4.$$

Let R_0 be the collection of all rational numbers. Under (A1), $n[\hat{\tau}_{1k}(x) - \tau_{1k}(x)] \to 0$ a.s., for any $x \in R_0$. Hence, almost surely,

$$n\left[\hat{\tau}_{1k}(x) - \tau_{1k}(x)\right] \to 0 \quad \text{for all } x \in R_0. \tag{4.11}$$

For fixed sequences $\{\hat{\tau}_{1k}(x)\}$ and $\{\hat{\theta}_k\}$ satisfying (4.11) and $\hat{\theta}_k - \theta \to 0$, select $\varepsilon > 0$ such that $\theta \pm \varepsilon \in R_0$ and $\theta - \varepsilon \le \hat{\theta}_k \le \theta + \varepsilon$ for large k. Since $\hat{\tau}_{1k}(x)$ is nondecreasing in x, $\hat{\tau}_{1k}(\theta - \varepsilon) \le \hat{\tau}_{1k}(\hat{\theta}_k) \le \hat{\tau}_{1k}(\theta + \varepsilon)$ and

$$\begin{aligned}
&\hat{\tau}_{1k}(\theta - \varepsilon) - \tau_{1k}(\theta - \varepsilon) + \tau_{1k}(\theta - \varepsilon) - \tau_{1k}(\theta) \\
&\quad \le \hat{\tau}_{1k}(\hat{\theta}_k) - \tau_{1k}(\theta) \\
&\quad \le \hat{\tau}_{1k}(\theta + \varepsilon) - \tau_{1k}(\theta + \varepsilon) + \tau_{1k}(\theta + \varepsilon) - \tau_{1k}(\theta).
\end{aligned}$$

Letting $k \to \infty$, we have

$$\begin{aligned}
\lim_{k\to\infty} n\left[\tau_{1k}(\theta - \varepsilon) - \tau_{1k}(\theta)\right] &\le \liminf_{k\to\infty} n\left[\hat{\tau}_{1k}(\hat{\theta}_k) - \tau_{1k}(\theta)\right] \\
&\le \limsup_{k\to\infty} n\left[\hat{\tau}_{1k}(\hat{\theta}_k) - \tau_{1k}(\theta)\right] \\
&\le \lim_{k\to\infty} n\left[\tau_{1k}(\theta + \varepsilon) - \tau_{1k}(\theta)\right].
\end{aligned} \tag{4.12}$$

From the definition of $\tau_{1k}(x)$ and (A1),

$$\begin{aligned}
&n\left[\tau_{1k}(\theta + \varepsilon) - \tau_{1k}(\theta)\right] \\
&= E\sum_{h=1}^{L} \frac{n(1 - f_h)n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[a_{hi}(\theta + \varepsilon) - a_{hi}(\theta)\right]\left[a_{hi}(\theta + \varepsilon) + a_{hi}(\theta)\right] \\
&\le 4n \max_{j \le N_{hi}, i \le N_h, h \le L} n_{hi} w_{hij} E \sum_{h=1}^{L} \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} w_{hij} I(\theta < y_{hij} \le \theta + \varepsilon) \\
&= O(1)E\left[G_k(\theta + \varepsilon) - G_k(\theta)\right] = O(1)\left[F_k(\theta + \varepsilon) - F_k(\theta)\right].
\end{aligned}$$

Hence by (A9), $\lim_{\varepsilon \to 0} \lim_{k \to \infty} n[\tau_{1k}(\theta + \varepsilon) - \tau_{1k}(\theta)] = 0$. This result still holds if ε is replaced by $-\varepsilon$. Letting $\varepsilon \to 0$ in (4.12), we have $\lim_{k \to \infty} n[\hat{\tau}_{1k}(\hat{\theta}_k) - \tau_{1k}(\theta_k)] = 0$. This (and the subsequence argument) proves that

$$n\left[\hat{\tau}_{1k}(\hat{\theta}_k) - \tau_{1k}(\theta_k)\right] \to_p 0.$$

Similarly, we can show that, for $l = 2, 3$ and 4,

$$n\left[\hat{\tau}_{lk}(\hat{\theta}_k) - \tau_{lk}(\theta_k)\right] \to_p 0 \quad \text{and} \quad \hat{F}_k(\hat{\theta}_k) - F_k(\theta_k) \to_p 0.$$

Hence (4.9) follows from

$$\hat{v}_k = \left[\hat{F}_k^2(\hat{\theta}_k)\widehat{\mathrm{var}}(\infty) - 2\hat{F}_k(\hat{\theta}_k)\left[\hat{\tau}_{3k}(\hat{\theta}_k) - \hat{\tau}_{4k}(\hat{\theta}_k)\right] + \hat{\tau}_{1k}(\hat{\theta}_k) - \hat{\tau}_{2k}(\hat{\theta}_k)\right] \Big/ q_k^2$$

and

$$v_k(\theta_k) = F_k^2(\theta_k)\,\mathrm{Var}(q_k) - 2F_k(\theta_k)\left[\tau_{3k}(\theta_k) - \tau_{4k}(\theta_k)\right] + \tau_{1k}(\theta_k) - \tau_{2k}(\theta_k).$$

This completes the proof. □

From results (4.4) and (4.10), an approximate $(1-2\alpha)$ confidence interval for θ_k is

$$CI_k = \left[\widehat{\theta}_k - z_\alpha \widehat{\sigma}_k,\ \widehat{\theta}_k + z_\alpha \widehat{\sigma}_k\right],$$

where z_α is the $(1-\alpha)$ quantile of the standard normal distribution (or the t-distribution with $n-1$ degrees of freedom). Another confidence interval is Woodruff's (1952) interval obtained by inverting the sample distribution function. Let $s_k = \left[v_k(\theta_k)\right]^{1/2}$. If s_k were known, then by (4.3), an approximate $(1-2\alpha)$ confidence interval for $\widehat{\theta}_k$ is

$$CI_k^* = \left[\widehat{F}_k^{-1}(p - z_\alpha s_k),\ \widehat{F}_k^{-1}(p + z_\alpha s_k)\right].$$

The "confidence interval" CI_k^* is better than CI_k since it avoids the estimation of $1/f_k(\theta_k)$. Since s_k is unknown, it has to be replaced by $\widehat{s}_k = (\widehat{v}_k)^{1/2}$, and the resulting confidence interval is Woodruff's interval:

$$CI_k^W = \left[\widehat{F}_k^{-1}(p - z_\alpha \widehat{s}_k),\ \widehat{F}_k^{-1}(p + z_\alpha \widehat{s}_k)\right].$$

Asymptotic validity of CI_k^W was first shown by Francisco and Fuller (1991), but they assumed some unnecessary conditions [Conditions 5–7 in Francisco and Fuller (1991)]. We now prove the same result with the conditions in Theorem 5.

THEOREM 6. *Suppose that assumptions* (A1) *and* (A6)–(A9) *hold. Then*

$$P\left\{\theta_k \in CI_k^W\right\} \to 1 - 2\alpha. \tag{4.13}$$

PROOF. From (4.9), $\widehat{s}_k / s_k \to_p 1$. Let $\varepsilon > 0$ be arbitrarily given. Then

$$P\left\{(1-\varepsilon)s_k \le \widehat{s}_k \le (1+\varepsilon)s_k\right\} \to 1. \tag{4.14}$$

Because (4.14), we may assume $(1-\varepsilon)s_k \le \widehat{s}_k \le (1+\varepsilon)s_k$ in the following proof. Note that

$$\begin{aligned} &P\left\{\theta_k \in \left[\widehat{F}_k^{-1}\left(p - z_\alpha(1-\varepsilon)s_k\right),\ \widehat{F}_k^{-1}\left(p + z_\alpha(1-\varepsilon)s_k\right)\right]\right\} \\ &\quad \le P\left\{\theta_k \in CI_k^W\right\} \\ &\quad \le P\left\{\theta_k \in \left[\widehat{F}_k^{-1}\left(p - z_\alpha(1+\varepsilon)s_k\right),\ \widehat{F}_k^{-1}\left(p + z_\alpha(1+\varepsilon)s_k\right)\right]\right\} \end{aligned}$$

and, by (4.3),

$$\begin{aligned} &P\left\{\theta_k \in \left[\widehat{F}_k^{-1}\left(p - z_\alpha(1\pm\varepsilon)s_k\right),\ \widehat{F}_k^{-1}\left(p + z_\alpha(1\pm\varepsilon)s_k\right)\right]\right\} \\ &= P\left\{\theta_k \le \widehat{F}_k^{-1}\left(p + z_\alpha(1\pm\varepsilon)s_k\right)\right\} - P\left\{\theta_k < \widehat{F}_k^{-1}\left(p - z_\alpha(1\pm\varepsilon)s_k\right)\right\} \\ &= P\left\{\theta_k \le \widehat{\theta}_k + z_\alpha(1\pm\varepsilon)s_k / \left[f_k(\theta_k)n^{1/2}\right] + o_p(n^{-1/2})\right\} \\ &\quad - P\left\{\theta_k < \widehat{\theta}_k - z_\alpha(1\pm\varepsilon)s_k / \left[f_k(\theta_k)n^{1/2}\right] + o_p(n^{-1/2})\right\} \\ &\to \quad 2\Phi\left(z_\alpha(1\pm\varepsilon)\right) - 1, \end{aligned}$$

where Φ is the standard normal distribution function. Hence

$$\begin{aligned} 2\Phi\big(z_\alpha(1-\varepsilon)\big) - 1 &\le \liminf_{k\to\infty} P\{\theta_k \in CI_k^W\} \\ &\le \limsup_{k\to\infty} P\{\theta_k \in CI_k^W\} \\ &\le 2\Phi\big(z_\alpha(1+\varepsilon)\big) - 1. \end{aligned}$$

The result follows since ε is arbitrary. □

The variance estimator proposed by Francisco and Fuller (1991) is

$$\widetilde{\sigma}_k^2 = \widehat{v}_k \widetilde{u}_k^2, \tag{4.15}$$

where

$$\widetilde{u}_k = (2z_\alpha \widehat{s}_k)^{-1}\big[F_k^{-1}(p + z_\alpha \widehat{s}_k) - F_k^{-1}(p - z_\alpha \widehat{s}_k)\big]. \tag{4.16}$$

This estimator is motivated by Woodruff's interval CI_k^W. Comparing (4.8) with (4.16), and using (4.3), we can show that $\widetilde{\sigma}_k^2/\widehat{\sigma}_k^2 \to_p 1$.

Acknowledgments. I would like to thank Professor J. N. K. Rao at Carleton University, two referees and an Associate Editor for their helpful comments.

REFERENCES

BEACH, C. M. and KALISKI, S. F. (1986). Lorenz curve inference with sample weights: an application to the distribution of unemployment experience. *J. Roy. Statist. Soc. Ser. C* **35** 38–45.

BICKEL, P. J. (1973). On some analogues to linear combinations of order statistics in the linear model. *Ann. Statist.* **1** 597–616.

BICKEL, P. J. and FREEDMAN, D. A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *Ann. Statist.* **12** 470–482.

CHERNOFF, H., GASTWIRTH, J. L. and JOHNS, M. V., JR. (1967). Asymptotic distribution of linear combinations of order statistics, with applications to estimation. *Ann. Math. Statist.* **38** 52–72.

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

FRANCISCO, C. A. and FULLER, W. A. (1991). Quantile estimation with a complex survey design. *Ann. Statist.* **19** 454–469.

KISH, L. and FRANKEL, M. R. (1974). Inference from complex samples (with discussion). *J. Roy. Statist. Soc. Ser. B* **36** 1–37.

KOENKER, R. and BASSETT, G. (1978). Regression quantiles. *Econometrica* **46** 33–50.

KOENKER, R. and PORTNOY, S. (1987). L-estimation for linear models. *J. Amer. Statist. Assoc.* **82** 851–857.

KREWSKI, D. and RAO, J. N. K. (1981). Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *Ann. Statist.* **9** 1010–1019.

MCCARTHY, P. J. (1969). Pseudo-replication: half samples. *Internat. Statist. Rev.* **37** 239–264.

NYGARD, F. and SANDSTRÖM, A. (1985). The estimation of Gini and the entropy inequality parameters in finite populations. *Journal of Official Statistics* **1** 399–412.

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.

RUPPERT, D. and CARROLL, R. J. (1980). Trimmed least squares estimation in the linear model. *J. Amer. Statist. Assoc.* **75** 828–838.

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

SHAO, J. and WU, C. F. J. (1989). A general theory for jackknife variance estimation. *Ann. Statist.* **17** 1176–1197.

SHAO, J. and WU, C. F. J. (1992). Asymptotic properties of the balanced repeated replication method for sample quantiles. *Ann. Statist.* **20** 1571–1593.

STIGLER, S. M. (1973). The asymptotic distribution of the trimmed mean. *Ann. Statist.* **1** 472–477.

TUKEY, J. (1958). Bias and confidence in not quite large samples (abstract). *Ann. Math. Statist.* **29** 614.

WELSH, A. H. (1987). The trimmed mean in the linear model (with discussion). *Ann. Statist.* **15** 20–45.

WOODRUFF, R. S. (1952). Confidence intervals for medians and other position measures. *J. Amer. Statist. Assoc.* **47** 635–646.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706-1885