PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 15 (1995), pp. 85–99

ASYMPTOTIC THEORY OF LINEAR STATISTICS IN SAMPLING PROPORTIONAL TO SIZE WITHOUT REPLACEMENT

BY

P. J. BICKEL* (BERKELEY, CALIFORNIA)
AND
V. N. NAIR** (ANN ARBOR, MICHIGAN)

Abstract. Consider an ordered sample that is selected from a finite population successively without replacement and with probability proportional to some measure of size. In this paper, we study the asymptotic behavior of linear statistics from such a sampling scheme. Unlike previous results in the literature which consider only order--invariant statistics, we study the asymptotic distribution of linear statistics that depend on the order in which the sample is observed. Such statistics arise in the course of studying the nonparametric maximum likelihood estimators of the finite population and of the unknown population size. The asymptotic behavior is studied under conditions that are weaker than those assumed previously, and we also obtain simpler proofs of some existing results.

1. Introduction. Let $U_1, \ldots, U_N$ denote a finite population of N units, and let $w_j > 0$ be a size measure associated with U_j, $j = 1, \ldots, N$. We obtain an ordered sample of n units from the finite population as follows. The first unit is selected randomly according to the selection probabilities $\tilde{w}_{1j}$'s, where $\tilde{w}_{1j} = w_j / \sum_{k=1}^{N} w_k$. The selected unit, say U_{k_1}, is removed from the population, and the next unit is selected from the remaining part of the population according to the selection probabilities $\tilde{w}_{2j}$'s, where $\tilde{w}_{2j} = w_j / (\sum_{r=1}^{N} w_r - w_{k_1})$. The procedure is repeated until the n units have been selected. The probability of observing the ordered sample $(U_{k_1}, \ldots, U_{k_n})$ is thus

$$(1.1)\qquad \prod_{i=1}^{n} \frac{w_{k_i}}{\sum_{j=1}^{N} w_j - \sum_{r=0}^{i-1} w_{k_r}}, \qquad \text{where } w_{k_0} \equiv 0.$$

* P. J. Bickel's research was supported in part by NSA Grant 904-94-H-2020.
** V. N. Nair's research was supported in part by NSF Grant DMS-94.043000.

This scheme is well known in survey sampling where it has been called variously as *ppswor* (*probability proportional to size without replacement*) *sampling* and *successive sampling with unequal probabilities.* In survey sampling, the w_j is an auxiliary variable associated with the population units, and the values of the w_j's are all known a priori. When the auxiliary variable is correlated with the underlying variable of interest, it is advantageous to use the auxiliary information to sample the units with unequal probabilities (see, e.g., [7], Chapter 5). Holst [8] considered a generalization of this scheme where, in addition, there is a fixed cost of sampling associated with each unit. The sampling scheme is to select units successively and stop as soon as the cost exceeds a fixed budget. Gordon [5] called this a VCSS (*variable cost successive sampling*) *scheme.*

More recently, the scheme in (1.1) has been used as a model for incorporating size-biases inherent in discovery data. In petroleum resource estimation, for example, it is well known that measures of size of a pool, such as area, impact its chance of discovery. Typically, the "larger" pools tend to be discovered earlier. To account for such bias, Kaufman and his colleagues have proposed modeling the discovery process as sampling successively from a finite population without replacement and with probability proportional to some measure of size. Barouch and Kaufman [2] considered the case where the size measure is the area of the oil and gas pools. Others have extended this to the case where w_j is area raised to some power θ. The parameter θ is called the *coefficient of discoverability* in the petroleum estimation literature, with a larger value of θ indicating a more efficient discovery process. Nair and Wang [9] considered a more general setup where $w_j = \prod_{k=1}^{K} y_{kj}^{\theta_k}$, where the y_{kj}'s are different attributes, such as area, mean formation depth and net pay thickness, that are likely to impact chance of discovery.

Note that, in the above case, the size measures are not known a priori and that the $(N-n)$ w_j's associated with the unobserved units remain unknown even after the sample is observed. Often, the size measures are the variables of study. Therefore, methods of estimation must rely on only the attributes based on the observed sample. In some applications, the population size N is also unknown.

Many authors have studied estimation issues related to the *ppswor* scheme in survey sampling. See Hedayat and Sinha [7] and references therein. Inference in situations where the w_j's are not known a priori and the estimation has to be based solely on the observed sample has also been studied by several authors. Most of these results were obtained under the assumption that a population characteristic such as the population size N or the population total of the characteristic of interest is known. See Nair and Wang [9] and the references therein for estimation under a superpopulation framework where the finite population itself is assumed to be an *iid* sample from some underlying population. Gordon [5] and Andreatta and Kaufman [1] discussed inference procedures for the finite population itself and considered, among other things, Horvitz–Thompson type estimators. Gordon [5] also provided a moment-type estimator of the population size using a split-sample technique. Bickel et al. [3] studied the nonparametric maximum likelihood estimator (NPMLE) of the finite population, including that of the population size N. In [3], the asymptotic behavior of the NPMLE was studied under the assumption that the population has only a finite number of distinct values. In this article, we develop asymptotic theory that would allow us to relax this assumption and extend the results in [3] to more general situations.

The asymptotic theory of *ppswor* sampling has been studied by several authors. Most of them (Rosén [10], Holst [8], Sen [11], Hájek [6], and Gordon [4]) developed their results under the stringent assumption that the size measures are bounded away from zero. This was relaxed by Gordon [5] who assumed only a moment condition. However, all of these authors studied only linear statistics of the form $\sum_{j=1}^{N} b_j e_{nj}$, where e_{nj} is the indicator of the event that U_j is selected in a sample of size n. These statistics are *order invariant*, i.e., they do not depend on the order in which the sample is observed. In studying the properties of the NPMLE, one has to deal with statistics that depend, in general, on the order information (see [3]). To consider such statistics, define, for $1 \leqslant i \leqslant n$,

$$e_{ij} = \mathit{1}(U_j \text{ is selected in ordered sample of size } i), \tag{1.2}$$

where $\mathit{1}(A)$ denotes the indicator of the event A. Then the statistics of interest are of the form

$$T = \sum_{i=1}^{n} \sum_{j=1}^{N} g_{ij}(e_{ij}),$$

where $g_{ij}(\cdot)$ is known.

We consider only linear statistics in this paper and derive their limiting distributions in Sections 2 and 3. We obtain the results under conditions that are weaker than those considered previously. Our results also simplify the proofs of some of the existing results in Gordon [5]. These results will be used elsewhere to establish the asymptotic properties of some nonparametric estimators under *ppswor* sampling for more general situations than those considered in [3].

2. Limit theorems for linear statistics. We first consider linear statistics of the form

$$T_n \equiv \sum_{i=1}^{n} \sum_{j=1}^{N} a_i b_j e_{ij}. \tag{2.1}$$

In the sequel, a_i, b_j, e_{ij}, w_j all depend on n and N but we will suppress the dependence for notational simplicity. We make the following assumptions. Here ε, δ, M and so on denote constants independent of n.

A1: If $N \equiv N_n$ and $f_n \equiv n/N$, then for some $\varepsilon > 0$ we have $\varepsilon \leqslant f_n \leqslant 1-\varepsilon$ for all n.

A2: If $w_j > 0$ is the size measure of U_j, then $N^{-1}\sum_{j=1}^{N} w_j^{2+2\delta} \leqslant M_1$ for all n for some $\delta > 0$ and $M_1 < \infty$.

A3: For every $\varepsilon > 0$, there exists $\delta > 0$ such that $N^{-1}\sum_{j=1}^{N} 1(w_j \leqslant \delta) \leqslant \varepsilon$ for all n.

A4: For all n,

$$\text{(i)} \sum_{j=1}^{N} b_j = 0; \quad \text{(ii)} \sum_{j=1}^{N} b_j^2 = 1; \quad \text{(iii)} \max_{1 \leqslant j \leqslant N} |b_j| \to 0.$$

A5: $\sum_{i=1}^{n} |a_i| \leqslant M_2$ for all n.

A6: Either

(i) $a_1 = \ldots = a_{[\delta n]} = 0$ for some $\delta > 0$, for all n

or

(ii) $\sum_{j=1}^{N} |b_j| w_j^2 = O(N^{1/2})$ and $\sum_{j=1}^{N} b_j^2 w_j^2 = O(1)$.

Note that A6 (ii) is satisfied if A2 holds and $\max |b_j| = O(N^{-1/2})$.

Throughout, let $\pi(t) = 1-e^{-t}$ and $\bar{\pi}(t) \equiv 1-\pi(t) = e^{-t}$. Define λ_i as the unique solution of

$$\sum_{j=1}^{N} \pi(\lambda w_j) = i, \quad 1 \leqslant i \leqslant N, \tag{2.2}$$

and $\lambda_0 \equiv 0$. Further, let

$$\mu_n = \sum_{j=1}^{N} \sum_{i=1}^{n} a_i b_j \pi(\lambda_i w_j). \tag{2.3}$$

Define

$$b_{ij} \equiv b_j - \bar{b}_i, \quad \text{where } \bar{b}_i = \frac{\sum_{j=1}^{N} b_j w_j \bar{\pi}(\lambda_i w_j)}{\sum_{j=1}^{N} w_j \bar{\pi}(\lambda_i w_j)},$$

and

$$\sigma_n^2 = \sum_{j=1}^{N} \Big\{ \sum_{i=1}^{n} (B_{ij} - \bar{B}_j)^2 \big(\bar{\pi}(\lambda_{i-1} w_j) - \bar{\pi}(\lambda_i w_j)\big) + \bar{B}_j^2 \bar{\pi}(\lambda_n w_j) \Big\}, \tag{2.4}$$

where

$$B_{ij} \equiv \sum_{k=i}^{n} a_k b_{kj} \quad \text{and} \quad \bar{B}_j \equiv \sum_{i=1}^{n} B_{ij} [\bar{\pi}(\lambda_{i-1} w_j) - \bar{\pi}(\lambda_i w_j)].$$

THEOREM 1. *Under* A1–A5:

(i) *If for some* $\delta > 0$, $\sigma_n^2 \geqslant \delta$ *for all* n, *then* $(T_n - \mu_n)/\sigma_n$ *tends in law to* $N(0, 1)$.

(ii) *If* $\sigma_n = o(1)$, *then* $T_n - \mu_n = o_p(1)$.

Note 1. Conditions A4 (i) and (ii) are not real restrictions. Given $\{b_j\}$, define

(2.5) $$\tilde{b}_j = \frac{(b_j - b.)}{\sum_{i=1}^{N} [(b_i - b.)^2]^{1/2}}.$$

Then $\{\tilde{b}_j\}$ satisfy A4 (i), (ii) and

$$\frac{T_n - \mu_n}{\sigma_n} = \tilde{\sigma}_n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N} a_i \tilde{b}_j (e_{ij} - \pi(\lambda_i w_j)),$$

where $\tilde{\sigma}_n^2$ is given by (2.4) with b_j replaced by $\tilde{b}_j$ in the definition of B_j's.

COROLLARY 1 (Gordon [5]). *Suppose* $a_1 = \ldots = a_n - 1 = 0$, $a_n = 1$ *so that*

$$T_n = \sum_{j=1}^{N} b_j e_{nj},$$

where $e_{nj} \equiv 1$ *(U_j included in sample of size n). Suppose that* A1–A3 *hold, and further that*

(2.6) $$N^{-1} \sum_{j=1}^{N} w_j^{-\gamma} \leqslant M_3$$

for all n, some $\gamma > 0$, $M_3 < \infty$. *Then the conclusion of Theorem* 1 *holds with*

$$\mu_n = \sum_{j=1}^{N} b_j \pi(\lambda_n w_j) \quad \textit{and} \quad \sigma_n^2 = \sum_{j=1}^{N} (b_j - \bar{b}_n)^2 \pi(\lambda_n w_j) \bar{\pi}(\lambda_n w_j).$$

Corollary 1 follows readily from Theorem 1 since

$$N^{-1} \sum_{j=1}^{N} 1(w_j \leqslant \delta) \leqslant \delta^\gamma N^{-1} \sum_{j=1}^{N} w_j^{-\gamma}.$$

Note 2. Gordon states his result (Theorem 2.2d) without the normalization on the b_j's. But it is not hard to see that, for the *psswor* sampling setup considered in this paper, his statistics and ours can be identified with a suitable definition of the b_j's. Gordon in fact develops his result for Holst's [8] generalization of *ppswor* sampling. We do not consider Holst's generalization here, but it is also not hard to see that our result can be extended to cover Gordon's result also in that situation.

We now extend Theorem 1 slightly to handle an important situation that arises in a statistical problem of interest. Consider

(2.7) $$\tilde{T}_n = N^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{N} (c_{ij}/N + b_j)(e_{ij} - \pi(\lambda_i w_j)).$$

Let X_{jN}'s be the attributes associated with the finite population units U_{jN}'s, and suppose the X_{jN}'s belong to a common Euclidean space X. Assume:

B0: $f_n \to f_0$, $0 < f_0 < 1$.

B1: If F_n is the empirical distribution function of the X_j's, suppose $F_n \Rightarrow F$ with $F\{0\} = 0$.

B2: There exist continuous functions $w_n(\cdot)$ and $w(\cdot)$: $X \to R^+$ such that
(i) $w_n(X_{jN}) = w_j$,
(ii) $w_n(x) \to w(x)$ uniformly on compacts, and
(iii) $\int w_n^{2+2\delta}(x)dF_n(x) \leqslant M < \infty$ for all n and some $\delta > 0$.

Evidently, B0–B2 imply A1–A3. Define λ_n: $[0, 1] \to R^+$ by

$$\int_X \bar{\pi}(\lambda_n w_n(x))dF_n(x) = 1 - s,$$

and define λ analogously with F_n replaced by F. From B1 and B2 we obtain

$$\int_X \bar{\pi}(\lambda_n w_n(x))dF_n(x) \to \int_X \bar{\pi}(\lambda w(x))dF(x)$$

uniformly on R^+. Hence

$$\lambda_n(s) \to \lambda(s) \tag{2.8}$$

and, by the analyticity of λ,

$$\lambda_n'(s) \to \lambda'(s) \tag{2.9}$$

uniformly in s for $\varepsilon \leqslant s \leqslant 1-\varepsilon$ for $\varepsilon > 0$.

B3: There exist continuous functions a_k: $[0, 1] \to R$, $1 \leqslant k \leqslant K$, b_k: $X \to R$, $0 \leqslant k \leqslant K$, such that
(i) $c_{ijn} = \sum_{k=1}^{K} a_{kn}(i/N)b_{kn}(X_{jN})$, $b_j = b_{0n}(X_{jN})$, and
(ii) $a_{kn}(s) \to a_k(s)$ uniformly on $[0, 1]$ for $1 \leqslant k \leqslant K$, and $b_{jn}(x) \to b_j(x)$ uniformly on compacts, $0 \leqslant j \leqslant K$.

B4: (i) $\int b_{jn}(x)dF_n(x) = 0$ for all n and $0 \leqslant j \leqslant K$,
(ii) $\varepsilon \leqslant \sum_{j=0}^{K} \int b_{jn}^{2+\delta}(x)dF_n(x) \leqslant O(1)$ for some δ, $\varepsilon > 0$, and
(iii) $\sup_x \sum_{j=0}^{K} |b_{jn}(x)| = O(N^{1/2})$.

B5: Either
(i) $a_{jk}(s) \equiv 0$ for $0 \leqslant s \leqslant \delta_k$ for some $\delta_k > 0$, $1 \leqslant k \leqslant K$ and all n,
or
(ii) $\int |b_{jn}(x)|\, w_n^2(x)dF_n(x) = O(1)$ for $0 \leqslant j \leqslant K$.

Let

$$B_j(s) = \int b_j(x)w(x)\bar{\pi}(\lambda(s)w(x))dF(x) / \int w(x)\bar{\pi}(\lambda(s)w(x))dF(x), \tag{2.10}$$

$$C(t, x) = \sum_{r=1}^{K} \int_t^{f_0} a_r(s)(b_r(x) - B_r(s))ds, \tag{2.11}$$

$$\bar{C}(x) = \sum_{r=1}^{K} b_r(x) \int_0^{f_0} a_r(s)\bar{\pi}(\lambda(s)w(x))w(x)\lambda'(s)ds, \tag{2.12}$$

$$(2.13)\qquad \bar{B}(x) = b_0(x)\big(\pi(\lambda(f_0)w(x))\big) - \int_0^{f_0} B_0(s)\bar{\pi}\big(\lambda(s)w(x)\big)w(x)\lambda'(s)ds,$$

and

$$(2.14)\qquad \sigma^2 = \int_X \Big\{ \int_0^{f_0} [C(t, x) - \bar{C}(x)]^2 w(x)\bar{\pi}\big(\lambda(s)w(x)\big)\lambda'(s)ds + [\bar{B}(x) + \bar{C}(x)]^2 \bar{\pi}\big(\lambda(f_0)w(x)\big)\Big\} dF(x).$$

THEOREM 2. *If* B0–B5 *hold, then* $\tilde{T}_n$ *given by* (2.7) *converges weakly to* $N(0, \sigma^2)$.

COROLLARY 2. *If* $\tilde{T}_{nj}$, $1 \leqslant j \leqslant J$, *are of the form* (2.7) *and satisfy* B0–B5, *then* $(\tilde{T}_{n1}, \ldots, \tilde{T}_{nJ})$ *converges weakly to a multivariate* $N(\boldsymbol{0}, \Sigma)$ *distribution where*

$$\Sigma = \Big\| \int [C_a - \bar{C}_a][C_b - \bar{C}_b]dv_1 + \int [\bar{C}_a + \bar{B}_a][\bar{C}_b + \bar{B}_b]dv_2 \Big\|,$$

C_a, C_b, *etc. are defined suitably and* v_1 *is the measure on* $\boldsymbol{X} \times [0, 1]$ *with density* $w(x)\bar{\pi}\big(\lambda(s)w(x)\big)\lambda'(s)$ *with respect to the product of* F *and the Lebesgue measure, and* v_2 *is the measure on* X *with density* $dv_2/dF = \bar{\pi}\big(\lambda(f_0)w(x)\big)$.

Note 3. Theorems 1 and 2 and the lemmas which imply them (see Section 3) parallel Propositions 1–4 and Corollary 1 of [3]. In fact, Corollary 1 generalizes Proposition 2 in [3] and Theorem 2 generalizes in part Proposition 4 and Corollary 1 in [3].

Note 4. We can extend Theorem 1 and the corollary to the general case where $c_{ij} = c_n(i/n, j/N)$; see Proposition 4 of [3]. But the present form is as general as we need for our inference problem of interest.

3. Proofs. Consider the following scheme. Let $N_1(t), \ldots, N_N(t)$, $t > 0$, be independent homogeneous Poisson processes with rates $w_1, \ldots, w_N$, respectively. Let τ_1 be the first time such that one of the $N_j(\cdot)$ jumps and let J_1 be the index j of the corresponding process. Correspondingly, let (τ_2, J_2) be the time and index of the next process to jump etc. Let $\tilde{e}_{ij} \equiv \boldsymbol{1}(J_k = j$ for some $k \leqslant i)$. It is shown in [3], following Rosén [10], that

$$(3.1)\qquad \{\tilde{e}_{ij}\},\ 1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant N, \quad \text{and} \quad \{e_{ij}\},\ 1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant N,$$

have the same joint distribution. Define the random time-scale transformation consisting of stopping times with respect to the filtration induced by $(N_1(\cdot), \ldots, N_N(\cdot))$,

$$(3.2)\qquad \tau_n(s) \equiv \begin{cases} \tau_i, & i/n \leqslant s < (i+1)/n,\ i \geqslant 1, \\ 0, & 0 \leqslant s < 1/n, \end{cases}$$

and, similarly,

$$\lambda_n(s) \equiv \begin{cases} \lambda_i, & i/n \leqslant s < (i+1)/n,\ i \geqslant 1, \\ 0, & 0 \leqslant s < 1/n. \end{cases} \tag{3.3}$$

If T_n is given by (2.1) and we identify $\{\tilde{e}_{ij}\}$ and $\{e_{ij}\}$, we can write

$$T_n - \mu_n = \sum_{i=1}^{n} a_i \sum_{j=1}^{N} b_j\big(1\big(N_j(\tau_n(i/n)) > 0\big) - \pi(\lambda_n(i/n)w_j)\big). \tag{3.4}$$

Let

$$\bar{b}(t) \equiv \Big(\sum_{j=1}^{N} b_j w_j \bar{\pi}(tw_j)\Big)\Big/\Big(\sum_{j=1}^{N} w_j \bar{\pi}(tw_j)\Big), \qquad b_j(t) \equiv b_j - \bar{b}(t). \tag{3.5}$$

Note that from (3.5) and the definition of τ_n and λ_n we obtain

$$T_n - \mu_n = \sum_{i=1}^{n} \sum_{j=1}^{N} b_j(\lambda_n(i/n))\big(1\big(N_j(\tau_n(i/n)) > 0\big) - \pi(\lambda_n(i/n)w_j)\big). \tag{3.6}$$

Let

$$N(t) = \sum_{j=1}^{N} N_j(t) \Big/ \sum_{j=1}^{N} w_j.$$

Note that $EN(t) = t$. Define

$$W_{n0}(t) = N^{-1/2} \sum_{j=1}^{N} \big(1(N_j(t) > 0) - \pi(tw_j)\big), \tag{3.7}$$

$$W_{n1}(t) = \sum_{j=1}^{N} b_j(t)\big(1(N_j(t) > 0) - \pi(tw_j)\big), \tag{3.8}$$

$$W_{n2}(t) = \sum_{j=1}^{N} b_j(\tau_n(t))\big(\pi(\tau_n(t)w_j) - \pi(\lambda_n(t)w_j)\big), \tag{3.9}$$

and let A_n place mass a_i at i/n, $1 \leqslant i \leqslant n$. Then, since

$$\sum_{j=1}^{N} \big(1\big(N_j(\tau_n(t)) > 0\big) - \pi(\lambda_n(t)w_j)\big) = 0 \quad \text{for } t = i/n,\ i \geqslant 0,$$

we get from (3.6)–(3.9)

$$T_n - \mu_n = \int_0^1 \big(W_{n1}(\tau_n(s)) + W_{n2}(s)\big)dA_n(s). \tag{3.10}$$

We need some preliminary lemmas. Let $\|g\| \equiv \sup\{|g(x)|\colon 0 \leqslant x \leqslant M\}$ for $M < \infty$ fixed, and for processes U_n, V_n on $[0, M]$ write

$$U_n(t) = V_n(t) + o_p(1) \quad \text{if and only if} \quad \|U_n - V_n\| = o_p(1).$$

We will use repeatedly the following elementary inequality:

LEMMA 1. *For all* $x,\ y > 0,\ 0 \leqslant \delta \leqslant 1$,

$$|e^{-y}-e^{-x}-(x-y)e^{-x}| \leqslant C|y-x|^{1+\delta}. \tag{3.11}$$

Proof. Let $\Delta = x-y$. Then (3.11) is equivalent to

$$|e^{\Delta}-1-\Delta| \leqslant Ce^{x}|\Delta|^{1+\delta}.$$

If $|\Delta| \leqslant 1$,

$$|e^{\Delta}-1-\Delta| \leqslant \frac{e}{2}\Delta^2$$

and we can take $C = e/2$. If $\Delta \geqslant 1$,

$$|e^{\Delta}-1-\Delta| \leqslant e^{\Delta} \leqslant e^{x} \leqslant e^{x}\Delta^{1+\delta},$$

while if $\Delta < -1$,

$$|e^{\Delta}-1-\Delta| \leqslant |\Delta|+2 \leqslant 3e^{x}|\Delta|^{1+\delta},$$

and the result follows.

LEMMA 2. *Under* A1–A4 *and* A6 (ii),

$$\tau_n(t) = \lambda_n(t)+o_p(1), \tag{3.12}$$

$$W_{n1}(\tau_n(t)) = W_{n1}(\lambda_n(t))+o_p(1). \tag{3.13}$$

If A1–A4 *and* A6 (i) *hold, then* (3.12) *and* (3.13) *are valid if* τ_n, λ_n, $W_{n1}(\tau_n)$, $W_{n1}(\lambda_n)$ *are replaced by* $\tau_n 1_A$, $\lambda_n 1_A$, $W_{n1}(\tau_n 1_A(t))$, $W_{n1}(\lambda_n 1_A(t))$, *where* $A = \{t:\ \varepsilon \leqslant t \leqslant M\}$, $\varepsilon > 0$.

Note 5. Claim (3.12) is aready proved in [10]. We include a proof here for completeness. In fact, we will prove a refined version of (3.12) in Lemma 3.

Proof of Lemma 2. We claim that the processes $W_{nj}(t)$, $j = 0, 1$, are:
(a) tight in $D[0, M]$, and
(b) have as possible limits only Gaussian processes with mean 0, the limiting covariance structure and continuous sample functions.

We give the argument for W_{n1} and compute for $s \leqslant t \leqslant u$

(3.14)
$$\mathrm{E}\big(W_{n1}(u)-W_{n1}(t)\big)^2\big(W_{n1}(t)-W_{n1}(s)\big)^2 \leqslant 2\Big(\sum_{j=1}^{N}\mathrm{E}\,V_{1j}V_{2j}\Big)^2+\sum_{j=1}^{N}\mathrm{E}\,V_{1j}^2\sum_{j=1}^{N}\mathrm{E}\,V_{2j}^2,$$

where

$$V_{1j} = b_j(u)\big(1(N_j(u) > 0)-\pi(w_j u)\big)-b_j(t)\big(1(N_j(t) > 0)-\pi(w_j t)\big)$$

and V_{2j} is defined similarly to W_{n2} in (3.9). A tedious computation gives

$$\mathrm{E}\,V_{1j}^2 = b_j^2(t)\pi(w_jt)\bar{\pi}(w_jt) - 2b_j(s)b_j(t)\bar{\pi}(w_jt)\pi(w_js) + b_j^2(s)\pi(w_js)\bar{\pi}(w_js),$$

$$\mathrm{E}\,V_{1j}V_{2j} = \big(b_j(u)\bar{\pi}(w_ju) - b_j(t)\bar{\pi}(w_jt)\big)\big(b_j(t)\pi(w_jt) - b_j(s)\pi(w_js)\big).$$

Now, for $0 \leqslant t \leqslant M$,

$$\frac{\partial b_j(t)}{\partial t} = -\frac{\sum_{k=1}^N b_k w_k^2 \bar{\pi}(tw_k)}{\sum_{k=1}^N w_k \bar{\pi}(tw_k)} + \bar{b}(t)\frac{\sum_{j=1}^N w_k^2 \bar{\pi}(tw_k)}{\sum_{j=1}^N w_k \bar{\pi}(tw_k)} \tag{3.15}$$

$$= O\Big(N^{-1}\sum_{k=1}^N |b_k| w_j^2 + N^{-1/2}\Big).$$

Since $\partial\pi(w_jt)/\partial t \leqslant w_j$, if A6 (ii) holds, we can easily deduce that, for $0 \leqslant s \leqslant t \leqslant M$,

$$\mathrm{E}V_{1j}^2 \leqslant C_1 \sup\{N^{-1/2}|b_j(v)| + w_j b_j^2(v)\colon\ 0 \leqslant v \leqslant M\}(t-s) \tag{3.16}$$

$$\leqslant C_2\{N^{-1/2}b_j + N^{-1} + w_j(b_j^2 + N^{-1})\}(t-s).$$

Similarly, since

$$\left|\frac{\partial}{\partial t}[b_j(t)\pi(w_jt)]\right| + \left|\frac{\partial}{\partial t}[b_j(t)\bar{\pi}(w_jt)]\right| \leqslant C_3\Big(N^{-1/2} + N^{-1}\sum_{k=1}^N |b_k| w_k^2 + N^{-1/2}w_j\Big),$$

we conclude that

$$|\mathrm{E}V_{1j}V_{2j}| \leqslant C_4\big(N^{-1}(1+w_j^2)\big) + \Big(N^{-1}\sum_{k=1}^N |b_k| w_k^2\Big)^2\big)(u-s)^2. \tag{3.17}$$

Note that if we consider only $0 < \varepsilon \leqslant s \leqslant t \leqslant M$, we can replace $N^{-1}\sum_{j=1}^N |b_j| w_j^2$ in (3.15) and (3.16) by $N^{-1}\sum_{j=1}^N |b_j| = O(N^{-1/2})$ and $w_j^2 b_j^2$ and $w_j b_j^2$ in (3.16), (3.17) by b_j^2. If A6 (ii) holds or we consider $W_{n1}(s)$ only on A_0, we conclude that (3.14) is bounded by $c(t-u)^2$, and (a) follows by the Billingsley–Centsov inequalities. Further, (b) follows from a similar but easier bound for $\mathrm{E}\big(W_{n1}(s) - W_{n1}(t)\big)^2$. A similar argument applies to $W_{n0}(\cdot)$.

It is now clear that (3.13) follows from (3.12) and (a) and (b). For (3.12) note that

$$N^{-1}\sum_{j=1}^N 1(N_j(t) > 0) = N^{-1}\sum_{j=1}^N \pi(w_jt) + o_p(1). \tag{3.18}$$

Since

$$N^{-1}\sum_{j=1}^N \pi(w_jt) \geqslant N^{-1}\sum_j \{\pi(\varepsilon t)\colon\ w_j \geqslant \varepsilon\} \quad \text{and} \quad \lim_{\varepsilon\to 0}\lim_n N^{-1}\sum_{j=1}^N 1(w_j < \varepsilon) = 0,$$

we conclude that for $0 \leqslant t \leqslant 1$ we have $0 \leqslant \lambda_n(t) \leqslant M < \infty$ for some M and all n. Now, (3.17) implies that

$$P[\tau_n(t) \leqslant 2M,\, 0 \leqslant t \leqslant 1] \to 1$$

and (3.12) follows from (3.18) and the monotonicity of $N^{-1}\sum_{j=1}^{N} 1(N_j(t) > 0)$ and $N^{-1}\sum_{j=1}^{N} \pi(w_j t)$. The lemma is proved.

LEMMA 3. *Under* A1–A4 *and* A6 (ii),

$$W_{n2}(t) = o_p(1). \tag{3.19}$$

If A1–A4 *hold, then* (3.19) *holds for* $W_{n2}(1_A(t))$.

Proof of Lemma 3. We refine (3.19) to

$$\pi_n(t)-\lambda_n(t) = -N^{-1/2}W_{n0}(\lambda_n(t))\Big(N^{-1}\sum_{j=1}^{N} w_j\Big)^{-1}(1+o_p(1)) = O_P(N^{-1/2}). \tag{3.20}$$

To see this, note

$$\begin{aligned}(3.21)\quad & N^{-1}\sum_{j=1}^{N}\big(\pi(w_j\tau_n(t))-\pi(w_j\lambda_n(t))\big)\\ &= \Big(N^{-1}\sum_{j=1}^{N} w_j\Big)\big(\tau_n(t)-\lambda_n(t)\big)+O_P\Big(N^{-1}\sum_{j=1}^{N} w_j^{1+\delta}|\tau_n(t)-\lambda_n(t)|^{1+\delta}\Big)\\ &= \Big(N^{-1}\sum_{j=1}^{N} w_j+o_p(1)\Big)\big(\tau_n(t)-\lambda_n(t)\big)\end{aligned}$$

by (3.12).

On the other hand,

$$\begin{aligned}(3.22)\quad N^{-1}\sum_{j=1}^{N}\big(\pi(w_j\tau_n(t))-\pi(w_j\lambda_n(t))\big) &= -N^{-1/2}W_{n0}(\tau_n(t))\\ &= -N^{-1/2}W_{n0}(\lambda_n(t))(1+o_p(1))\end{aligned}$$

by (3.12) again, and (3.20) follows. Now

$$\begin{aligned}(3.23)\quad |W_{n2}(t)| &= \Big|\sum_{j=1}^{N} b_j(\tau_n(t))\big[\pi(w_j\lambda_n(t))-\pi(w_j\tau_n(t))-w_j\bar{\pi}(w_j\tau_n(t))\big]\Big|\\ &\leqslant \sum_{j=1}^{N} |w_j|^{1+\delta}|b_j(\tau_n(t))|\,|\tau_n(t)-\lambda_n(t)|^{1+\delta}\end{aligned}$$

by Lemma 1.

The right-hand side of (3.23) is

$$O_P\Big(\Big(\sum_{j=1}^{N} |w_j|^{2+2\delta}\Big)^{1/2} N^{-1/2-\delta/2}\Big) = o_p(1)$$

by A2 and (3.21). The lemma follows.

Proof of Theorem 1. From (3.11)–(3.13) and (3.19) we obtain

(3.24)

$$W_{n1}(\tau_n(t)) + W_{n2}(t) = \sum_{j=1}^{N} b_j(\lambda_j(t))\big[1\big(N_j(\lambda_n(t) > 0) - \pi(\lambda_n(t) w_j)\big)\big] + o_p(1)$$

under A1–A4 and A6 (ii). Then, from (3.10) and A5 we get

$$T_n - \mu_n = \sum_{j=1}^{N} \sum_{i=1}^{n} a_i b_j(\lambda_i)\big[1\big(N_j(\lambda_i) > 0\big) - \pi(\lambda_i w_j)\big] + o_p(1). \tag{3.25}$$

Evidently, (3.25) still holds if A6 (ii) is replaced by A6 (i). Theorem 1 now follows from the Lindeberg–Feller theorem. To see this note first that, by A3,

$$\max_j \Big|\sum_{i=1}^{n} a_i b_j(\lambda_i)\Big| = O\big(\max_{i,j} |b_j(\lambda_i)|\big) = O\big(\max |b_j| + \max |\bar{b}(\lambda_i)|\big) \tag{3.26}$$

$$= O(\max |b_j|) = o(1)$$

by A4 (iii). Further,

(3.27)

$$\operatorname{var}\Big(\sum_{j=1}^{N} \sum_{i=1}^{n} a_i b_j(\lambda_i) 1\big(N_j(\lambda_i) > 0\big)\Big) = \sum_{j=1}^{N} \operatorname{var}\Big(\sum_{i=1}^{n} a_i b_j(\lambda_i) 1\big(N_j(\lambda_i) > 0\big)\Big) = \sigma_n^2$$

and the theorem follows.

Proof of Theorem 2. To simplify the notation, we take $K = 1$. If we let, for $k = 0, 1$,

$$\tilde{b}_{kn} = (b_{kn} - \bar{b}_k)/\tau_{kn}, \quad \text{where } \tau_{kn}^2 = N^{-1} \sum_{j=1}^{N} b_{kn}^2(X_{jN}),$$

we see that

$$\tilde{T}_n = \tilde{T}_{n1} + \tilde{T}_{n0}, \tag{3.28}$$

where

$$\tilde{T}_{n1} = N^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N} a_n(i/N) \tilde{b}_{1n}(X_{jN})\big(e_{ij} - \pi(\lambda_i w_i)\big),$$

and

$$T_{n0} = \sum_{i=1}^{n} \sum_{j=1}^{N} \tilde{b}_{0n}(X_{jN})(e_{ij} - \pi(\lambda_i w_j)).$$

By the proof of Theorem 1, if $\{a_n\}$, $\{b_{1n}\}$, and $\{b_{0n}\}$ satisfy A4–A6, then

(3.29)

$$\tilde{T}_n = \sum_{i=1}^{n} \sum_{j=1}^{N} (N^{-1} a_n(i/N) b_{1j}^*(\lambda_i) + b_{0j}^*(\lambda_i))(I(N_j(\lambda_i) > 0) - \pi(\lambda_i w_j)) + o_P(1),$$

where, with an abuse of notation, we define

$$b_{kj}^*(\lambda) = \tilde{b}_{kn}(X_{jN}) - \bar{b}_{kn}(\lambda), \tag{3.30}$$

and

$$\bar{b}_{kn}(\lambda) = \frac{\sum_{r=1}^{N} b_{kn}(X_{rN}) w_r \bar{\pi}(\lambda w_r)}{\sum_{r=1}^{N} w_r \bar{\pi}(\lambda w_r)}. \tag{3.31}$$

To prove the theorem, we, therefore, need only check that $\{a_n\}$, $\{\tilde{b}_{1n}\}$, and $\{\tilde{b}_{0n}\}$ satisfy A4–A6 and that the variance of the leading term in (3.30) tends to σ^2. In view of B4 (i), the $\tilde{b}_{kn}$ satisfy A4 (i) and (ii). But

$$\max_r |\tilde{b}_{kn}(X_{rN})| = N^{-1/2} O(\max_r |B_{kn}(X_{rN})|) = o(1),$$

by B4 (ii) and (iii). Hence A4 (iii) follows. A5 and A6 are immediate consequences of B3 (ii) and B5.

Let

$$B_{kn}(s) = \frac{\int b_{kn}(x) w_n(x) \bar{\pi}(\lambda(s) w_n(x)) dF_n(x)}{\int w_n(x) \bar{\pi}(\lambda(s) w_n(x)) dF_n(x)}. \tag{3.32}$$

Then $B_{kn}(i/n) = \bar{b}_{kn}(\lambda_i)$. Let v_n be the distribution that assigns mass $1/N$ to i/N, $1 \leqslant i \leqslant N$. Define

$$C_n(t, x) = \int_t^{f_n} a_n(s)[b_{1n}(x) - B_{1n}(s)] dv_n(s), \tag{3.33}$$

$$\bar{C}_n(x) = b_{1n}(x) \int_0^{f_n - 1/N} a_n(s)[\bar{\pi}(\lambda_n(s) w_n(x)) - \bar{\pi}(\lambda_n(s + 1/N) w_n(x))] N dv_n(s) \tag{3.34}$$

and

$$\bar{B}_n(x) = b_{0n}(x)[\pi(\lambda_n(f_n) w_n(x))] - \int_0^{f_n - 1/N} B_{0n}(s)[\bar{\pi}(\lambda_n(s) w_n(x)) - \bar{\pi}(\lambda_n(s + 1/N) w_n(x))] N dv_n(s). \tag{3.35}$$

Arguing as in (3.27), we can show

(3.36)

$$\sigma_n^2 \equiv \operatorname{var} \sum_{j=1}^{N} \Big(\sum_{i=1}^{n} [N^{-1} a_n(i/N) b_{1j}^*(\lambda_i) + b_{0j}^*(\lambda_i)] 1(N_j(\lambda_i) > 0) \Big)$$

$$= \int \Big[\int_0^{f_n - 1/N} [C_n(t, x) - \bar{C}_n(x)]^2 [\bar{\pi}(\lambda_n(s) w_n(x)) - \bar{\pi}(\lambda_n(s+1/N) w_n(x))] N dv_n(s)$$

$$+ [\bar{B}_n(x) + \bar{C}_n(x)]^2 \bar{\pi}(\lambda_n(f_n) w_n(x)) \Big] dF_n(x).$$

But, by B1–B4 (uniform integrability, convergence of integrands, and weak convergence of F_n),

(3.37) $$B_{kn}(s) \to B_k(s) \text{ uniformly on } [0, 1].$$

Similarly, B0–B4 and (3.37) imply

(3.38) $$C_n(t, x) \to C(t, x) \quad \text{for all } (t, x).$$

Further, using (2.9), we obtain

(3.39) $$\bar{C}_n(x) \to \bar{C}(x),$$

and

(3.40) $$\bar{B}_n(x) \to \bar{B}(x).$$

Finally, B4 gives the uniform integrability needed to conclude that σ_n^2 given by (3.36) converges to σ^2 given by (2.14), completing the proof.

REFERENCES

[1] G. Andreatta and G. M. Kaufman, *Estimation of finite population properties when sampling is without replacement and proportional to magnitude*, J. Amer. Statist. Assoc. 81 (1986), pp. 657–666.

[2] E. Barouch and G. M. Kaufman, *Probabilistic modelling of oil and gas discovery*, in: *Energy: Mathematics and Models*, F. Roberts (Ed.), SIAM, Philadelphia, 1976, pp. 248–260.

[3] P. J. Bickel, V. N. Nair and P. C. C. Wang, *Nonparametric inference under biased sampling from a finite population*, Ann. Statist. 20 (1992), pp. 853–878.

[4] L. Gordon, *Successive sampling in large finite populations*, Ann. Statist. 11 (1983), pp. 702–706.

[5] – *Estimation for large successive samples with unknown inclusion probabilities*, Adv. in Appl. Math. 14 (1993), pp. 89–112.

[6] J. Hájek, *Sampling from a Finite Population*, Marcel Dekker, New York 1981.

[7] A. S. Hedayat and B. K. Sinha, *Design and Inference in Finite Population Sampling*, Wiley, New York 1991.

[8] L. Holst, *Some limit theorems with applications in sampling theory*, Ann. Statist. 1 (1973), pp. 644–658.

[9] V. N. Nair and P. C. C. Wang, *Maximum likelihood estimation under a successive sampling discovery model*, Technometrics 31 (1989), pp. 423–436.

[10] B. Rosén, *Asymptotic theory for successive sampling with varying probabilities without replacement, I* and *II*, Ann. Math. Statist. 43 (1972), pp. 373–397, and pp. 748–776.

[11] P. K. Sen, *Invariance principles for the coupon collector's problem: A martingale approach*, Ann. Statist. 7 (1979), pp. 372–380.

P. J. Bickel
University of California, Berkeley
Berkeley, CA 94720, U.S.A.

V. N. Nair
University of Michigan
Ann Arbor, MI 48109, U.S.A.
and
AT&T Bell Laboratories
Murray Hill, NJ 07974, U.S.A.

Received on 2.4.1994