The Annals of Statistics
1983, Vol. 11, No. 3, 959–969

ASYMPTOTIC THEORY OF SYSTEMATIC SAMPLING

BY RONALDO IACHAN

Iowa State University

The main purpose of this paper is the development of an asymptotic theory of systematic sampling from a stochastic population. The superpopulation model assumed is that the population arises from a second-order stationary process. A comparison among the multiple random start systematic sampling schemes is made in terms of the limiting expected variance of the sample mean. Asymptotic normality of the systematic sampling mean is obtained, both unconditionally and conditionally on the given population. Finally, the asymptotic behavior of confidence intervals based on two distinct variance estimators is studied.

1. Introduction. Suppose we are interested in estimating the mean $\bar{X}_N$ of a finite population $\Pi_N = \{X(1), \cdots, X(N)\}$ with a natural labeling. Consider systematic samples of size n consisting of elements $X(j), X(j+k), \cdots, X(j+k(n-1))$ where j is a random integer between 1 and k. These are called single random start systematic samples, or, simply, systematic samples. Multiple random start systematic samples consist of elements indexed by m (say) random starts drawn without replacement from the first $\ell = mk$ labels and of every ℓth element thereafter.

Our asymptotic framework consists of a sequence Π_N of populations expanding as $N \to \infty$, from which samples of size $n_N = N/k \to \infty$ as $N \to \infty$ are taken. See, for example, Fuller and Isaki (1982) for a careful construction of a related framework. Except for Theorem 3.2, where k as well as n is allowed to grow to infinity with the population size, we shall assume k to be fixed throughout this work. This assumption could actually be relaxed by assuming only that $k_N \to k < \infty$ as $N \to \infty$, but the gain in generality in our results would be then overshadowed by the increased complexity of statements and proofs.

In order to study the asymptotic properties of the systematic sampling procedures as the sample size $n \to \infty$, we shall assume that the population itself is stochastic. Stochastic population (or superpopulation) models were introduced by Cochran (1946) in the very context of systematic sampling to formalize the strong dependence of the behavior of this sampling scheme on the structure of the population. More specifically, Cochran (1946) assumed that the population values themselves arise as the realization of a weakly stationary process.

If we introduce the usual p-notation for the sampling distribution and the suffix p for expectations over repeated applications of the sampling procedure, then for any design for which the sample mean $\bar{X}(n)$ is p-unbiased we have

$$V_p(\bar{X}(n)) = E_p(\bar{X}(n) - \bar{X}_N)^2.$$

In order to compare the precision of distinct (p-unbiased) sampling designs, Cochran (1946) argued that the p-variance $V_p(\bar{X}(n))$ itself cannot be used as a criterion since it depends heavily on the particular population at hand. Such was the motivation to examine classes of population within a superpopulation model and to consider instead $\mathscr{E}\{V_p(\bar{X}(n))\}$ as a criterion, where $\mathscr{E}$ denotes expectation under the superpopulation distribution. Note that the overall variance is

$$V(\bar{X}(n)) = \mathscr{E}\{V_p(\bar{X}(n))\} + \mathscr{V}\{E_p(\bar{X}(n))\},$$

Received February 1981; revised December 1982.

AMS 1980 *subject classification.* Primary 62D05; secondary 62F25, 62E20.

Key words and phrases. Multiple random start systematic sampling, limiting expected variance, asymptotic normality, variance estimator.

where $\mathscr{V}(Y) = \mathscr{E}(Y - \mathscr{E}Y)^2$ denotes variance with respect to the model distribution. The second term in the right hand side is however independent of the (p-unbiased) design and vanishes asymptotically. Our approach is thus distinct from that of Mátern (1960) who considers $n \times V_p(\bar{X}(n) - \bar{X}_N)$. We shall suppress hereon the suffix p when no confusion may arise.

It will be assumed that the population values are observations at $t = 1, 2, \cdots, N$ of a weakly (or second order) stationary process $\varepsilon(\cdot)$ defined on a certain probability space (Ω, F, P). Explicitly,

$$\mathscr{E}(\varepsilon(t)) = 0, \quad \mathscr{E}(\varepsilon(t)\varepsilon(t+v)) = \rho(v)\sigma^2, \quad t, v \in R. \tag{1.1}$$

We assume $\rho(0) = 1$ so σ^2 is the variance and $\rho(\cdot)$ the autocorrelation function of $\varepsilon(t)$.

A deterministic mean function was also considered in the model, and found to have no influence upon our asymptotic results provided that the mean is locally smooth. Its presence in the model, however, permits the study in the same framework of classical cases of departure from stationarity such as linear or parabolic trends (see Bellhouse and Rao, 1975, for example) and periodic components in the fixed finite population values. In case a linear trend is suspected, Yates' (1948) end corrections bring about a considerable improvement in estimation. Anyhow, it was shown by Madow (1953) in this case that centered systematic sampling (where the start is no longer random but located at the center of the first block of k elements) outperforms its random start counterpart.

If $\rho(\cdot)$ is non-negative, nonincreasing and convex, then as shown by Cochran (1946),

$$\mathscr{E}[V(\bar{X}_{\text{sys}}(n))] \leq \mathscr{E}[V(\bar{X}_{\text{strs}}(n))] \leq \mathscr{E}[V(\bar{X}_{\text{srs}}(n))] \tag{1.2}$$

where $\bar{X}_{\text{sys}}(n)$, $\bar{X}_{\text{strs}}(n)$ and $\bar{X}_{\text{srs}}(n)$ denote the means of systematic sampling, stratified random sampling (with one selection from each stratum of k consecutive elements) and simple random sampling respectively, all with sample size n. This result was extended by Gautschi (1957) and by Hájek (1959). In Section 2 we will partially generalize (exclusive of stratification) in an asymptotic setup all three results above by relaxing the assumptions on the model and by establishing a complete hierarchy of the multiple random start systematic sampling schemes in terms of their asymptotic expected variances. This result gives an idea of the cost (in terms of larger expected variance of the sample mean) incurred in using more random starts in order to afford a better estimate of the variance from a single sample. We also show in passing that under even weaker conditions, systematic sampling dominates simple random sampling. These results are discussed in more detail in Iachan (1980a).

We shall confine our study to designs in one dimension. Bellhouse (1977) showed that there can be no optimal unrestricted fixed sample size design in two dimensions even when the correlation function is assumed convex. He showed, however, that in this case systematic sampling has smallest expected variance in the restricted classes of aligned and unaligned designs introduced by Quenouille (1949). We should also remark that no loss of generality is incurred in assuming $N = nk$ for the sake of simplicity—the general case ($N \neq nk$) is handled similarly by means of Lahiri's circular procedure (see for example Murthy, 1967) and end effects are of no importance asymptotically.

Section 3 is devoted to the asymptotic normality of the systematic sampling mean, both unconditionally and conditionally on the given population, the latter with probability one, under some simple mixing conditions in model (1.1).

Section 4 deals with two variance estimators motivated by certain assumptions on $\rho(\cdot)$ and the asymptotic behavior of confidence intervals based on them. Section 5 contains the proofs of some of our results.

We now introduce some notations to be used throughout this work. Let $X_{ij} = X(j + k(i-1))$, $i = 1, \cdots, n$, $j = 1, \cdots, k$, denote the ith element of the jth systematic sample. Define the jth systematic sample mean $X_{\cdot j} = (1/n)\sum_{i=1}^{n} X_{ij}$ and let $\bar{X}_{\text{sys}}(n)$ be the random variable that takes on each $X_{\cdot j}$ value with probability $1/k$. Define in the same way $\bar{X}_{\text{sys}(m)}(n)$, the mean of systematic sampling with m random starts and sample size $n = mp$.

Let $V_{\text{sys}(m)}(n) = \mathscr{E}\{V(\bar{X}_{\text{sys}(m)}(n))\}$ and $V_{\text{sys}}(n) = V_{\text{sys}(1)}(n)$. The population mean can be written as $\bar{X}_N = (1/N) \sum_i \sum_j X_{ij}$.

2. A hierarchy of the multiple random start systematic sampling schemes. The correlation function $\rho(\cdot)$ is non-negative definite and even—we may and will consider it for $t \geq 0$ only. Let us impose our first assumption on the correlation function, a mild form of mixing condition on the process $\varepsilon(\cdot)$ in (1.1):

(A1) $\sum_{v=1}^{\infty} |\rho(v)| < \infty$.

It follows from (A1) that

(P1) $\lim_{t\to\infty}\rho(t) = 0$,

(P2) $\sum_{v=1}^{\infty} |\rho(kv)| < \infty$, $k \geq 1$, k integer.

Therefore, the function $S(k) = \sum_{v=1}^{\infty} \rho(kv)$ is well-defined, provided that (A1) is assumed as it will be from this point on.

We are now able to compute the limiting expected variance (normalized by a factor n) of systematic sampling from population (1.1). Here as everywhere in this work, $\ell = mk$ is the sampling interval (or step) of systematic sampling with m random starts and sampling fraction $1/k$.

THEOREM 2.1. *If* (A1) *is assumed in* (1.1) *then, as* $N \to \infty$ *and* $k = N/n$ *is held fixed,*

$$\lim_{n\to\infty} n V_{\text{sys}(m)}(n) = \sigma^2\left(1 - \frac{1}{k}\right)\left\{1 - \frac{2}{\ell - 1}(S(1) - \ell S(\ell))\right\} = V_{\text{sys}}(m), \quad \textit{say}, \tag{2.1}$$

where $S(k) = \sum_{v=1}^{\infty} \rho(kv)$.

It is easy to show that

$$V_{\text{srs}}(n) = \mathscr{E}\{\operatorname{Var}(\bar{X}_{\text{srs}}(n))\} = \frac{\sigma^2}{n}\left(1 - \frac{1}{k}\right)\left\{1 - \frac{2}{N(N-1)}\sum_{v=1}^{N-1}(N - v)\rho(v)\right\}$$

and

$$\lim_{n\to\infty} n V_{\text{srs}}(n) = \sigma^2\left(1 - \frac{1}{k}\right) = V_{\text{srs}}, \quad \text{say.}$$

The remainder of this section will be focused on the comparison of the multiple random start systematic sampling schemes—notice that both single random start systematic sampling and simple random sampling belong to this class (with $m = 1$ and $m = n$, respectively). Our ultimate goal here is a complete hierarchy of such sampling schemes in terms of their limiting expected variances

$$V_{\text{sys}}(m) = V_{\text{srs}}\left\{1 - \frac{2}{mk - 1}(S(1) - mkS(mk))\right\} = V_{\text{srs}}\{1 - 2g(mk)\}, \tag{2.2}$$

where we define the function $g(\ell) = (S(1) - \ell S(\ell))/(\ell - 1)$ for integers $\ell \geq 2$.

The comparison (2.2) is to be made between sampling schemes with the same sampling fraction $1/k$, hence k is to be kept fixed, and the comparison boils down to that between the values of $g(mk)$ for distinct numbers m of random starts, $1 \leq m \leq n$.

We will need in the next results some further assumptions on the correlation function:

(A2) $\rho(\cdot)$ is nonincreasing;

(A3) $\rho(\cdot)$ is convex;

(A4) $\rho(\cdot)$ is continuously differentiable on $(0, \infty)$ and $\sum_{v=1}^{\infty} |v\rho'(v)| < \infty$.

Note that we have imbedded our process in continuous time so that $\rho(\cdot)$, $S(\cdot)$ and $g(\cdot)$ are all defined in the real line.

THEOREM 2.2. *If the population in model* (1.1) *satisfies* (A1) *and* (A2) *then*

$$V_{\text{sys}}(m) \leq V_{\text{srs}}, \quad \textit{all } m \geq 1. \tag{2.3}$$

This theorem shows that systematic sampling with any number of random starts is better (in our usual sense) than simple random sampling provided that the correlation sequence $(\rho(v): v = 0, 1, 2, \cdots)$ is absolutely summable and nonincreasing. It is worth remarking that convexity of $\rho(\cdot)$ is no longer necessary for the asymptotic result, unlike the analogous exact results of Hájek (1959) and Gautschi (1957). However, the full set of assumptions (A1) to (A4) is necessary to establish a complete hierarchy of the multiple random start systematic sampling schemes, i.e.

$$V_{\text{sys}}(m_1) \leq V_{\text{sys}}(m_2) \quad \text{if} \quad m_1 \leq m_2. \tag{2.4}$$

Define the function $Q(\ell) = \ell S(\ell)$ and notice that $Q(\ell)$ and $S'(\ell) = \sum_v v\rho'(\ell v)$ are well-defined in view of (A1) and (A4), and hence so is $Q'(\ell)$. Our next result gives conditions for the hierarchy (2.4) to hold.

THEOREM 2.3. *If the population in model* (1.1) *satisfies* (A1) *to* (A4) *then*:
(i) *A necessary and sufficient condition for the hierarchy result* (2.4) *is that* $Q'(\ell) \geq \{Q(\ell) - Q(1)\}/(\ell - 1)$, *all* $\ell > 1$;
(ii) *A sufficient condition for* (2.4) *is the convexity of* $Q(\cdot)$.

It is not difficult to verify that the sufficient condition (ii) holds for the exponential correlation function $\rho(t) = e^{-\lambda t}$, $\lambda > 0$, but some relationship between ℓ and L is necessary if (ii) is to hold for the linear correlation function $\rho(t) = (1 - t/L)^+$, $L > 0$. Both are classical examples of concave upwards correlation functions, i.e., satisfying (A2) and (A3).

3. Asymptotic normality of the systematic sample mean. In order to establish asymptotic normality of the mean of systematic sampling (with sample size n, sampling fraction $1/k$) from population (1.1), we first need to find sufficient conditions for

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon(i) \to_{\mathscr{D}} \mathscr{N}(0, \tau^2) \quad \text{as} \quad n \to \infty, \tag{3.1}$$
$$\tau^2 = \sigma^2 \sum_{v=-\infty}^{\infty} \rho(v) = \sigma^2(1 + 2\sum_{v=1}^{\infty} \rho(v)).$$

This is done in the lemma that follows, which makes use of the following notation: a process $W(\cdot)$ is said to have all joint cumulants absolutely summable if for all integers $k \geq 1$, $\sum_{t_1} \cdots \sum_{t_k} |\text{cum}(W(t_1), \cdots, W(t_k))| < \infty$, the sum ranging over all integers. Recall that the joint cumulant of a k-dimensional random vector $(Y_1, \cdots, Y_k)$ can be defined as the coefficient of $i^k s_1 \cdots s_k$ in the Taylor expansion of $\log(E \exp(i \sum_{j=1}^{k} s_j Y_j))$.

LEMMA 3.1. *Either of the following two conditions is sufficient for* (3.1) *to hold*:
(B1) $\varepsilon(\cdot)$ *has finite moments of all orders and all joint cumulants absolutely summable*;
(B2) $\varepsilon(\cdot)$ *is a linear process*: $\varepsilon(t) = \sum_{i=-\infty}^{\infty} a_i Z_{t-i}$, *with* $\sum_i |a_i| < \infty$ *and* $(Z_n: n = 0, \pm 1, \cdots)$ *a sequence of* i.i.d. *random variables*, $\mathscr{E} Z_1 = 0$, $\mathscr{E} Z_1^2 < \infty$.

The proof that (3.1) holds for processes satisfying (B1) can be found in Brillinger (1975, page 94), and for processes in class (B2) the result follows from a theorem in Anderson (1971, page 429). We could have imposed milder mixing conditions, such as the strong mixing condition introduced by Rosenblatt (1956), or the ϕ-mixing condition detailed by Rosenblatt (1972), for example, to get stronger results along the same line. We chose, however, not to list alternative sets of conditions because the ones given encompass most of practical applications. Note, for instance, that Gaussian processes satisfying the summability assumption (A1) fall in class (B1).

We are now in a position to state our (unconditional) asymptotic normality result. Details of the proof can be found in Iachan (1980b).

THEOREM 3.1. *If in the population model* (1.1), *the process* $\varepsilon(\cdot)$ *satisfies* (3.1), *then as* $n \to \infty$ *and* $k = N/n$ *is held fixed,*

$$\sqrt{n}(\bar{X}_{\text{sys}}(n) - \bar{X}_N) \to_D \mathcal{N}(0, \sigma^2_{\text{sys}}), \sigma^2_{\text{sys}} = \sigma^2\left(1 - \frac{1}{k}\right)\left(1 + \frac{2}{k-1}(kS(k) - S(1))\right)$$
$$= \lim_{n\to\infty} n\mathscr{E}\{V(\bar{X}_{\text{sys}}(n))\} = V_{\text{sys}} \quad (\textit{see } (2.1) \textit{ or } (2.2)).$$

The following corollary is a trivial consequence of Lemma 3.1 and the subsequent remark.

COROLLARY 3.1. *If the process* $\varepsilon(\cdot)$ *satisfies either* (B1) *or* (B2) (*or any of the alternative mixing conditions referred to*) *then the asymptotic normality result* (3.1) *holds.*

An analogous result is valid for multiple random start systematic sampling. In case m random starts are employed, the asymptotic variance coincides with the limiting expected variance (2.1). The reader interested in the proof is referred to Iachan (1980b).

We turn now to asymptotic normality results conditional on the particular population at hand.

Note that under our assumptions, $\bar{X}_{\text{sys}}(n) - \bar{X}_N$ is the sum of components from n stretches of the series of population values generated by a process without long-range dependence. Let now $k \to \infty$ as well as $n \to \infty$, and define the sequence of distribution functions

$$F_{k,n}(x, w) = \Pr\{\sqrt{n}(\bar{X}_{\text{sys}}(n) - \bar{X}_N) \le x \mid X_1(w), \cdots, X_N(w)\} \tag{3.2}$$

for a given realization of population values

$$X_1(w), \cdots, X_N(w), \quad w \in \Omega.$$

THEOREM 3.2. *If the process* $\varepsilon(\cdot)$ *satisfies* (B1) *or* (B2) *then, there is a set* $A \subset \Omega$ *with* $P(A) = 1$ *such that,* $\forall\, w \in A$,

$$F_{k,n}(x, w) \to \Phi\left(\frac{x}{\sigma}\right), \quad \forall\, x, \tag{3.3}$$

as $k \to \infty$ *and* $n \to \infty$.

Here as elsewhere, Φ is the standard normal c.d.f., and P is the underlying distribution of process (1.1).

A detailed proof of this result (based on the convergence of the associated characteristic functions) can be found in Iachan (1980b) and is sketched in Section 5.

4. Variance estimators. The variance of systematic sampling cannot be unbiasedly estimated from a single sample. Various superpopulation models have been assumed in the literature to arrive at reasonable estimators—we will examine here the asymptotic behavior of two such estimators motivated by different assumptions on our model (1.1). Recall that P denotes the underlying distribution in superpopulation model (1.1) under which the systematic sampling variance estimators are to be studied.

The first variance estimator considered is based on the assumption that the population is in random order, translated into a correlation function in (1.1) which is zero if $t \neq 0$. The second estimator of the variance was introduced by Cochran (1946) motivated by a correlation function of the exponential form,

$$\rho(t) = \exp(-\lambda t), \quad t \ge 0, \lambda > 0. \tag{4.1}$$

Explicitly, if the jth systematic sample

$$(X_{ij} = X(j + k(i - 1)) \colon i = 1, \cdots, n)$$

is drawn, we form the estimators

$$\hat{V}_r(n) = \left(1 - \frac{1}{k}\right)\frac{1}{n}\textstyle\sum_{i=1}^{n} (X_{ij} - X_{\cdot j})^2/(n - 1) = \left(1 - \frac{1}{k}\right)W/n, \tag{4.2}$$

$$\hat{V}_c(n) = \left(1 - \frac{1}{k}\right)\frac{W}{n}\left(1 - \frac{2}{\ell n r_k^{-1}} + \frac{2}{r_k^{-1} - 1}\right), \tag{4.3}$$

where $W = W(j, n)$ is the mean of squares within the jth sample and $r_k = r_k(j, n) = \sum_{i=1}^{n-1} (X_{ij} - X_{\cdot j})(X_{i+1,j} - X_{\cdot j})/\sum_{i=1}^{n}(X_{ij} - X_{\cdot j})^2$ is a sample estimate of the autocorrelation at lag k, $\rho(k)$.

Consistency of r_k and of W (as an estimator of σ^2) as stated in Lemma 5.3 follows under a stronger assumption on model (1.1):

(C1) $\varepsilon(\cdot)$ is a stationary finite linear process of the (ARMA) form

$$\textstyle\sum_{i=1}^{p} a_i\varepsilon(t - i) = \sum_{j=1}^{q} b_j Z(t - j),$$

where $(Z(n) \colon n = 0, \pm 1, \cdots)$ is a sequence of i.i.d. mean zero random variables with a finite fourth cumulant.

Note that if a process $\varepsilon(\cdot)$ is in subclass (C1) of linear processes then it is also in (B2).

The theorem that follows provides a heuristic justification for the use of Cochran's estimator when the correlation function is exponential (as in (4.1)) and k as well as n is large.

THEOREM 4.1. *Under assumptions* (A1) *and* (C1) *on model* (1.1), *we have as* $n \to \infty$ *and* $k = N/n$ *is fixed*,

$$n\hat{V}_c(n) \to_P \sigma^2\left(1 - \frac{1}{k}\right)\left(1 - \frac{2}{\ell n \rho_k^{-1}} + \frac{2}{\rho_k^{-1} - 1}\right). \tag{4.4}$$

If we define now $c(k, \lambda) = P - \lim_{n\to\infty} n(\hat{V}_c(n) - V_{\text{sys}}(n))$ then it is not difficult to show that $\lim_{k\to\infty} c(k, \lambda) = 0$ for any $\lambda > 0$ in case the correlation function is indeed given by (4.1). See Iachan (1980c) for details.

The theorem also allows us to find an explicit expression for the following ratio which will be useful in future calculations:

$$R(k; \rho) = P - \lim_{n\to\infty}(\hat{V}_c(n)/V_{\text{sys}}(n)) \tag{4.5}$$

which in view of (2.1) and (4.4) equals

$$R(k; \rho) = \left(1 - \frac{2}{\ell n \rho_k^{-1}} + \frac{2}{\rho_k^{-1} - 1}\right)\Big/\left(1 - \frac{2}{k - 1}(S(1) - kS(k))\right). \tag{4.6}$$

The next two theorems deals with the asymptotic behavior of confidence intervals based on the two given variance estimators, namely intervals (with confidence level $1 - 2\alpha$) of the form

$$\bar{X}_{\text{sys}}(n) \pm Z_\alpha\sqrt{\hat{V}_r(n)}, \tag{4.7}$$

$$\bar{X}_{\text{sys}}(n) \pm Z_\alpha\sqrt{\hat{V}_c(n)}, \tag{4.8}$$

where Z_α is the upper α-point of the standard normal distribution.

THEOREM 4.2. *If* $\varepsilon(\cdot)$ *satisfies* (3.1) *and* (A2) *in model* (1.1), *then confidence intervals of the form* (4.7) *are asymptotically* ($n \to \infty$; $k = N/n$ *fixed*) *conservative.*

We can hope for more stringent attained levels (coverage probabilities) for intervals of the form (4.8). The next theorem shows, however, that these levels can be far too stringent even if the correlation function is nonincreasing convex.

THEOREM 4.3. *If in model* (1.1), *the process* $\varepsilon(\cdot)$ *satisfies* (3.1), (A2) *and* (A3), *then the coverage probabilities of intervals* (4.8) *can be asymptotically* ($n \to \infty$; *k fixed*) *zero.*

In order to be protected against the disastrous asymptotic performance of Cochran's estimator depicted by Theorem 4.3, we can correct it for asymptotic bias to obtain the estimator

$$\hat{V}_{MC}(n) = (1 - 1/k)\frac{W}{n}\left(1 - \frac{2}{k(r_k^{-1/k} - 1)} + \frac{2}{r_k^{-1} - 1}\right),$$

which we will call Modified Cochran's estimator and the associated confidence intervals

$$\bar{X}_{\text{sys}}(n) \pm Z_\alpha \sqrt{\hat{V}_{MC}(n)}. \tag{4.9}$$

We will establish an asymptotic lower bound for the coverage probability of such intervals. The theorem that follows thus provides an insurance against poor behavior of intervals (4.9).

THEOREM 4.4. *If in model* (1.1), *the process* $\varepsilon(\cdot)$ *satisfies* (3.1) *and* (A2), *then the lowest possible asymptotic* ($n \to \infty$; *k fixed*) *confidence level attained by intervals* (4.9) *is* $2\Phi(Z_\alpha/\sqrt{k}) - 1$.

We conclude that: (i) the estimator based on the assumption that $\rho(t) = 0$ if $t \neq 0$ in model (1.1)—the so called "random population" model—yields confidence intervals that are conservative asymptotically; (ii) the estimator based on the assumption that $\rho(t) = e^{-\lambda|t|}$, $\lambda > 0$, yields confidence intervals that can have zero asymptotic coverage probability if the assumption is wrong.

5. Proofs of some results.

5.1. *Proofs in Section 2.*

PROOF OF THEOREM 2.1. Consider first the case $m = 1$, i.e. the single random start systematic sampling expected variance $V_{\text{sys}}(n)$ which equals

$$\begin{aligned}\mathscr{E}\{V(\bar{X}_{\text{sys}}(n))\} &= \left\{\frac{1}{k}\sum_{j=1}^{j}(e_{\cdot j} - \bar{\varepsilon}_N)^2\right\} \\ &= \frac{\sigma^2}{n}\left(1 - \frac{1}{k}\right)\left\{1 - \frac{2}{N(k-1)}\sum_{v=1}^{N-1}(n - v)\rho(v)\right. \\ &\qquad \left. + \frac{2k}{n(k-1)}\sum_{v=1}^{n-1}(n - v)\rho(kv)\right\}.\end{aligned}$$

By a variation of Toeplitz Lemma, given e.g. in Heilbron (1968),

$$\sum_{v=1}^{N-1}\left(1 - \frac{v}{N}\right)\rho(v) \to \sum_{v=1}^{\infty}\rho(v) = S(1) \text{ as } N \to \infty, \sum_{v=1}^{N-1}\left(1 - \frac{v}{n}\right)\rho(kv)$$

$$\to \sum_{v=1}^{\infty}\rho(kv) = S(k) \text{ as } n \to \infty,$$

and hence

$$\lim_{n\to\infty} n V_{\text{sys}}(n) = \sigma^2\left(1 - \frac{1}{k}\right)\left\{1 - \frac{2}{k-1}S(1) + \frac{2k}{k-1}S(k)\right\} = V_{\text{sys}} \tag{5.1}$$

which proves (2.1) in this case. In order to generalize (5.1) for multiple random start systematic sampling, we notice first that

$$V\{\bar{X}_{\text{sys}(m)}(n)\} = c_{\ell,m}\frac{1}{\ell m}\sum_{j=1}^{\ell}(X_{\cdot j} - \bar{X}_N)^2,$$

where $c_{\ell,m} = (\ell - m)/(\ell - 1) = m(k-1)/(mk-1)$ is the finite population correction for sampling (without replacement) m out of ℓ elements, because $\bar{X}_{\text{sys}(m)} = (1/m) \cdot\sum_{r=1}^{m} X_{\cdot j_r}$, $i \le j_1 < j_2 < \cdots < j_m \le \ell$ being the random starts of the constituent systematic samples of size p, say. Therefore,

$$V\{\bar{X}_{\text{sys}(m)}(n)\} = \left(\frac{k-1}{\ell-1}\right)\sum_{j=1}^{\ell}(X_{\cdot j} - \bar{X}_N)^2/\ell = \left(\frac{k-1}{\ell-1}\right)V(\bar{X}_{\text{sys}}(p))$$

which yields, in view of (5.1),

$$\lim_{p\to\infty} p\{V(\bar{X}_{\text{sys}(m)}(n))\} = \left(\frac{k-1}{\ell-1}\right)\sigma^2\left(1-\frac{1}{\ell}\right)\left\{1 - \frac{2}{\ell-1}(S(1) - \ell S(\ell))\right\}$$

and (2.1) follows since $\ell = mk$ and $n = mp$. □

Proof of Theorem 2.2. In view of (2.2), all we need to show is that for all $\ell > 1$, $g(\ell) \ge 0$, or equivalently $S(1) \ge \ell S(\ell)$. But this follows by adding the inequalities

$$\rho(1) + \cdots + \rho(\ell) \ge \ell\rho(\ell)$$

$$\rho(\ell+1) + \cdots + \rho(2\ell) \ge \ell\rho(2\ell)$$

$$\cdots \quad \cdots \quad \cdots \quad \cdots$$

$$\rho(\ell(M-1)+1) + \cdots + \rho(\ell M) \ge \ell\rho(\ell M),$$

which hold for an arbitrary integer $M \ge 1$, and by letting $M \to \infty$. □

Proof of Theorem 2.3. (i) is equivalent to the non-positivity of the function

$$\frac{Q(\ell) - Q(1)}{\ell - 1} - Q'(\ell) = (\ell - 1)g'(\ell)$$

under the stated assumptions, and hence to the claimed monotonicity of $g(\cdot)$ for $\ell > 1$. (ii) follows from (i) and the fact that if $Q(\cdot)$ is convex then

$$-(\ell-1)Q'(\ell) \le \int_1^{\ell} - Q'(t)\,dt = Q(1) - Q(\ell),$$

where the inequality is due to the consequent monotonicity of $Q'(\cdot)$. □

5.2 *Proofs in Section 3.*

Sketch of the Proof of Theorem 3.2. From the unconditional result, it follows that

$$\frac{1}{k}\sum_{j=1}^{k}\mathscr{E}\exp(i\sqrt{n}X'_{\cdot j}t) \to \exp(-\sigma^2t^2/2), \quad \text{all} \quad t,$$

where we define

$$X'_{\cdot j} = X_{\cdot j} - \bar{X}_N$$

and script letters denote operators over the superpopulation distribution.

All we have to show is that, as $k, n \to \infty$,

$$\mathscr{V}\left\{\frac{1}{k}\sum_{j=1}^{k}\exp(i\sqrt{n}X'_{\cdot j}t)\right\}\to 0,\quad \text{all } t,$$

or equivalently, for any fixed value of t,

$$\frac{1}{k^2}\sum_{1\le \ell, m\le k} a_{n,k}(\ell, m)\to 0 \tag{5.2}$$

with $a_{n,k}(\ell, m) = \text{Cov}\{\exp(i\sqrt{n}X'_{\cdot\ell}t),\ \exp(i\sqrt{n}X'_{\cdot m}t)\}$. We need two lemmas in order to prove (5.2)

LEMMA 5.1. *Let $\ell = \ell_k = [ks]$, $m = m_k = [kw]$, $k = 1, 2, \cdots$, where s and w are distinct numbers in the unit interval. Under either* (B1) *or* (B2),

$$(\sqrt{n}\varepsilon_{\cdot\ell_k}, \sqrt{n}_{\cdot m_k})\to_D \mathscr{N}(0,\sigma^2)\times\mathscr{N}(0,\sigma^2)$$

i.e., a bivariate normal distribution with covariance matrix $\begin{pmatrix}\sigma^2 & 0\\ 0 & \sigma^2\end{pmatrix}$.

LEMMA 5.2. *Let $b_{n,k}(s, w) = a_{n,k}(\ell_k, m_k)$, $0\le s, w\le 1$. Under the conditions of the theorem, if $s\ne w$ then $b_{n,k}(s, w)\to 0$ pointwise as $k, n\to\infty$.*

Lemma 5.1 follows from convenient double array versions of the CLT for dependent variables. These Central Limit Theorems, as detailed in Iachan (1980b), are extensions of results in Diananda (1953) and Brillinger (1975, page 94). The method required a uniformity argument based on Anderson (1971, page 425).

Lemma 5.2 follows from the joint asymptotic normality and independence of $Y_{n,k} = \sqrt{n}X'_{\cdot\ell_k}$ and $Z_{n,k} = \sqrt{n}X'_{\cdot m_k}$, since in this case $\text{Cov}(g(Y_{n,k}), g(Z_{n,k}))\to 0$ as $k, n\to\infty$ for any bounded continuous $g(\cdot)$. Put $g(x) = e^{itx}$ to get $b_{n,k}\to 0$.

Finally, note that

$$\frac{1}{k^2}\sum_{1\le\ell, m\le k} a_{n,k}(\ell_k, m_k) = \int_0^1\int_0^1 b_{n,k}(s, w)\,ds\,dw + O(1/k)$$

and that the RHS $\to 0$ as $k, n\to\infty$ (by the DCT), since $b_{n,k}\to_{\text{a.e.}} 0$ in the unit square by Lemma 5.2. □

5.3 *Proofs in Section 4.*

PROOF OF THEOREM 4.1.

LEMMA 5.3. *Under assumptions* (A1), (C1) *and* (C2), *we have as $n\to\infty$ and $k = N/n$ is held fixed*:

(i) *W is a consistent estimator of σ^2;*

(ii) *r_k is a consistent estimator of $\rho_k = \rho(k)$.*

The proof of (i) follows from covariance formulae given, for example, by Hannan (1960), and (ii) is a consequence of Bartlett's (1946) formula (cf. Section 3.7 of Grenander and Rosenblatt, 1957) applied to the process $Y(t) = \varepsilon(j + kt)$.

A theorem attributed to Slutsky (cf. Billingsley, 1968, page 31) easily implies the following.

LEMMA 5.4. *Under the same assumptions, as $n\to\infty$ and $k = N/n$ is fixed, $\ell n\, r_k\to_P \ell n\rho_k$ and $1/r_k\to_P 1/\rho_k$.*

Theorem 4.1 is then a trivial consequence of the two lemmas and of the definition (4.3) of Cochran's estimator of the variance.

PROOF OF THEOREM 4.2. We shall first need a lemma.

LEMMA 5.5. *$\hat{V}_r$ is an asymptotic ($n \to \infty$; $k = N/n$ fixed) overestimate of the variance of systematic sampling from a population following model* (1.1) *with a nonincreasing correlation function $\rho(\cdot)$: $\hat{V}_r(n) \geq V_{\text{sys}}(n)$ with probability* 1 *for n large enough.*

PROOF. Recall from (2.2) that

$$nV_{\text{sys}}(n) \to_P \sigma^2\left(1 - \frac{1}{k}\right)(1 - 2g(k)) = V_{\text{sys}}, \tag{5.3}$$

with $g(k) = (S(1) - kS(k))/(k - 1)$. We have shown in the course of the proof of Theorem 2.2 that $g(k) \geq 0$, $k > 1$, if $\rho(\cdot)$ is nonincreasing. On the other hand, in view of Lemma 5.3, $n\hat{V}_r(n) \to_P \sigma^2(1 - 1/k)$, together with (5.3) yield the result. □

Under the given assumptions, we can apply Theorem 3.1 to get $\sqrt{n}(\bar{X}_{\text{sys}}(n) - \bar{\bar{X}}_N)/\sigma_{\text{sys}} \to_D \mathcal{N}(0, 1)$, so that, in view of (5.3) and Slutsky's Theorem, as $n \to \infty$ and k is fixed,

$$(\bar{X}_{\text{sys}}(n) - \bar{\bar{X}}_N)/\sqrt{V_{\text{sys}}(n)} \to_D \mathcal{N}(0, 1). \tag{5.4}$$

The coverage probability of interval (4.7) is given by

$$P\left\{\frac{|\bar{X}_{\text{sys}}(n) - \bar{X}_N|}{\sqrt{\hat{V}_r(n)}} \leq Z_\alpha\right\} \geq P\left\{\frac{|\bar{X}_{\text{sys}}(n) - \bar{X}_N|}{\sqrt{V_{\text{sys}}(n)}} \leq Z_\alpha\right\} \tag{5.5}$$

for large enough n, in view of Lemma 5.5. Since by (5.4) the right-hand side of (5.5) converges to $1 - 2\alpha$ (the nominal confidence level of intervals (4.7)), the theorem is proved. □

PROOF OF THEOREM 4.3. We will make use of the following lemma, the proof of which can be found in Iachan (1980c).

Let $H(x) = (\ln(1/x))^{-1} - (1/x - 1)^{-1}$, $0 < x < 1$.

LEMMA 5.6. *$H(x) \to 0$ as $x \to 0$, $H(x) \to ½$ as $x \to \infty$ and $H(\cdot)$ is nondecreasing in* [0, 1].

If we recall the definition (4.6) of the ratio $R(k; \rho)$ then, by a reasoning similar to the one in the proof of Theorem 4.2, all we need to prove is that $\lim\inf_{n\to\infty} R(k; \rho_n^*) = 0$ for some sequence (ρ_n^*) of nonincreasing convex correlation functions. We show that

$$\rho_n^*(t) = \left(1 - \frac{t}{n}\right)^+, \quad n = kp + r > k, \quad 0 < r < k,$$

satisfy the requirements. In view of Lemma 5.6, when k is fixed, $H(\rho_n^*(k)) \to ½$ as $n \to \infty$. Also, $\sum_{v=1}^{\infty} \rho_n^*(v) - k\sum_{v=1}^{\infty} \rho_n^*(kv) \to (k - 1 - r)/2$ as $n \to \infty$. Therefore,

$$\lim\inf_{n\to\infty} R(k; \rho_n^*) = \lim_{n\to\infty} \frac{1 - 2H(\rho_n^*(k))}{1 - 2\left(\dfrac{k - 1 - r}{2}\right)} = 0. \quad \square$$

PROOF OF THEOREM 4.4. We first state a lemma.

LEMMA 5.7. *Let $H_k(x) = 1/k(x^{-1/k} - 1) - 1(x^{-1} - 1)$. Then $H_k(x) \leq ½\,(1 - 1/k)$, $0 \leq x \leq 1$.*

For a proof of this lemma, we again refer to Iachan (1980c). If we define $R_{MC}(k; \rho) = P - \lim_{n\to\infty} n\hat{V}_{MC}(n)/V_{\text{sys}}$, the ratio necessary for our confidence interval calculations in this case, then it is easy to show as before that

$$R_{MC}(k;\rho) = \frac{1 - 2H_k(\rho(k))}{1 - 2g_k(\rho)}, \tag{5.6}$$

where we define the functional

$$g_k(\rho) = \{\textstyle\sum_{v=1}^{\infty} \rho(v) - k \sum_{v=1}^{\infty} \rho(kv)\}/(k-1).$$

Recall again from the proof of Theorem 2.2 that $g_k(\rho) \geq 0$ if ρ satisfies (A2), and then using Lemma 5.7 in (5.6) we obtain $R_{MC}(k;\rho) \geq 1/k$. Therefore, the result follows from

$$\lim_{n\to\infty} P\left\{\frac{|\bar{X}_{\text{sys}}(n) - \bar{X}_N|}{\sqrt{\hat{V}_{MC}(n)}} \leq Z_\alpha\right\} = 2\phi(Z_\alpha\sqrt{R_{MC}(k;\rho)}) - 1 \geq 2\phi(Z_\alpha/\sqrt{k}) - 1. \qquad \square$$

Acknowledgments. This work is based on part of the author's Ph.D. dissertation at the University of California at Berkeley under the guidance of Professor Peter J. Bickel, to whom the author is most grateful. The author is also thankful to the referees and to an Associate Editor for some helpful suggestions.

REFERENCES

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley, New York.

BARTLETT, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time series. *J. Roy Statist. Soc.* **8** 27–41.

BELLHOUSE, D. R. (1977). Some optimal designs for sampling in two dimensions. *Biometrika* **64** 605–611.

BELLHOUSE, D. R. and RAO, J. N. K. (1975). Systematic sampling in the presence of a trend. *Biometrika* **62** 694–697.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

BRILLINGER, D. R. (1975). *Time Series—Data Analysis and Theory.* Holt, New York.

COCHRAN, W. G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. *Ann. Math. Statist.* **17** 1964–77.

DIANANDA, P. H. (1953). Probability limit theorems with statistical applications. *Proc. Cambridge Philos. Soc.* **49** 239–246.

FULLER, W. and ISAKI, C. (1982). Survey design under the regression superpopulation model. *J. Amer. Statist. Soc.* **77** 89–96.

GAUTSCHI, W. (1957). Some remarks on systematic sampling. *Ann. Math. Statist.* **28** 385–394.

GRENANDER, U. and ROSENBLATT, M. (1957). *Statistical Analysis of Time Series.* Wiley, New York.

HÁJEK, J. (1959). Optimum strategy and other problems in probability sampling. *Casopis pro Pestovani Matematiky* **84** 387–420.

HANNAN, E. J. (1960). *Time Series Analysis.* Methuen, London.

HEILBRON, D. C. (1968). Estimating the variability of systematic sampling. Ph.D. Thesis, University of California, Berkeley.

IACHAN, R. (1980a). Comparison of multiple random start systematic sampling schemes. *J. Amer. Statist. Soc.*, to appear.

IACHAN, R. (1980b). An asymptotic theory of systematic sampling II. University of Wisconsin Tech. Report 617.

IACHAN, R. (1980c). An asymptotic theory of systematic sampling III. University of Wisconsin Tech. Report 618.

MADOW, W. G. (1953). On the theory of systematic sampling III: Comparison of centered and random start systematic sampling. *Ann. Math. Statist.* **24** 101–106.

MATÉRN, B. (1960). *Spatial Variation.* Meddelanden fran Statens Skogsforskningsinstitut 49:5.

MURTHY, M. N. (1967). *Sampling Theory and Methods.* Statistical Publishing Society, Calcutta.

QUENOUILLE, M. H. (1949). Problems in plane sampling. *Ann. Math. Statist.* **20** 355–375.

ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci.* **42** 43–47.

ROSENBLATT, M. (1972). Central limit theorem for stationary processes. *6th Berkeley Symposium on Probab. Statist.* **2** 551–561.

YATES, F. (1948). Systematic sampling. *Phil. Trans. Roy. Soc.* (*A*) **241** 347.

DEPARTMENT OF STATISTICS
102 SNEDECOR HALL
IOWA STATE UNIVERSITY
AMES, IOWA 50011