The Annals of Statistics
1990, Vol. 18, No. 4, 1615–1633

ACCURATE MULTIVARIATE ESTIMATION USING TRIPLE SAMPLING[1]

BY SHARON L. LOHR

University of Minnesota

Any multiresponse estimation experiment requires a decision about the number of observations to be taken. If the covariance is unknown, no fixed-sample-size procedure can guarantee that the joint confidence region will have an assigned shape and level. Double-sampling procedures use a preliminary sample of size m to determine the minimum number of additional observations needed to achieve a prescribed accuracy and coverage probability for the parameter estimates. The triple-sampling procedures of this paper, less sensitive to the choice of m, revise the sample size estimate after collecting a fraction of the additional observations prescribed under double sampling. Second-order asymptotic results relying on conditional inference show that triple sampling is asymptotically consistent; in addition, the regret for triple sampling is a bounded function of the covariance structure and is independent of m.

1. Introduction. Let $\mathbf{X}_1, \mathbf{X}_2, \dots$ be a sequence of independent and identically distributed random p-vectors with unknown mean $\boldsymbol{\theta}$ and unknown positive definite covariance matrix $\boldsymbol{\Sigma}$. The problem addressed in this paper is that of determining a sample size τ such that the resulting estimator $\hat{\boldsymbol{\theta}}_\tau$ accurately estimates $\boldsymbol{\theta}$. Accurate estimation is used here in the sense of Finster (1985, 1986). A fixed-accuracy set is a natural extension of a fixed-width confidence interval to $\mathscr{R}^p$: $\hat{\boldsymbol{\theta}}$ accurately estimates $\boldsymbol{\theta}$ with accuracy A and confidence γ if $P(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \in A) \geq \gamma$. Formally, a fixed-accuracy set is a compact, orientable Borel-measurable set $A \in \mathscr{R}^p$ which is star-shaped with respect to $\mathbf{0}$ and contains $\mathbf{0}$ as an interior point. The requirement that A be star-shaped ensures that if $\hat{\boldsymbol{\theta}}$ accurately estimates $\boldsymbol{\theta}$, so does any estimate $\tilde{\boldsymbol{\theta}}$ between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}$.

Accurate estimates are useful in a wide variety of applications. Often experimenters want a confidence region for a multivariate response which is of a specified shape and size and is easy to interpret. For example, the U.S. Environmental Protection Agency guidelines for solid waste analysis [U.S. Environmental Protection Agency (1982), page 5] state that it is desirable to use as few samples as necessary to achieve, with 80% confidence, a target joint accuracy in which the log concentration of the ith contaminant is estimated to

Received March 1988; revised September 1989.

[1]Research partially supported by University of Wisconsin-Madison Graduate School research grant 135–1001.

AMS 1980 *subject classifications*. Primary 62H12; secondary 62F25, 62L12.

Key words and phrases. Three-stage estimation, confidence sets, multivariate normal distribution, sequential methods.

within error d_i. In other words, their goal is a fixed-size rectangular accuracy region $A = \Pi[-d_i, d_i]$, rather than Working and Hotelling's (1929) ellipsoidal confidence set whose size and orientation depend upon the unknown covariance matrix. Fishman (1977) and Kleijnen (1984) describe the problem of determining the sample size to estimate the steady-state means of responses in a queueing simulation study. The procedures developed in this paper provide an algorithm for calculating the sample size and estimator for multiresponse computer simulation studies.

Dantzig (1940) showed that a sequential or step-sequential procedure is necessary to obtain accurate estimators in the multivariate normal situation with unknown covariance. In many cases, however, a purely sequential procedure, in which the parameters are reestimated after each observation, is impractical. Following Stein (1945), Cox (1952) and Hall (1981), who studied the one-dimensional case of fixed-width confidence interval estimation, we limit data collection to stages.

If $\boldsymbol{\Sigma}$ were known and the population were normal, any sample size n larger than $N = N(\boldsymbol{\Sigma})$, where $N(\mathbf{V})$ is the solution to

$$f(N(\mathbf{V}), \mathbf{V}) = \gamma \tag{1.1}$$

and

$$f(n, \mathbf{V}) = \int_A (n/2\pi)^{p/2} |\mathbf{V}|^{-1/2} \exp\left[-(n/2)\mathbf{x}^T \mathbf{V}^{-1}\mathbf{x}\right] d\mathbf{x}, \tag{1.2}$$

would ensure that $\overline{\mathbf{X}}_n$ is an accurate estimator of $\boldsymbol{\theta}$. For $\boldsymbol{\Sigma}$ unknown, double- and triple-sampling procedures both prescribe collecting a first sample of size m and estimating $\boldsymbol{\Sigma}$ by

$$\hat{\boldsymbol{\Sigma}}_m \equiv \frac{1}{m-1} \sum_{i=1}^{m} (\mathbf{X}_i - \overline{\mathbf{X}}_m)(\mathbf{X}_i - \overline{\mathbf{X}}_m)^T.$$

A natural estimator of N after the pilot sample has been collected is $\hat{N} = N(\hat{\boldsymbol{\Sigma}}_m)$. Lohr (1988) shows that $\hat{N}$ is asymptotically unbiased but the coverage probability using $\hat{N}$ is strictly less than γ because only a fraction of the data are used to estimate $\boldsymbol{\Sigma}$: the conditional distribution of $\hat{\boldsymbol{\Sigma}}_m^{-1/2}\overline{\mathbf{X}}$ (the normal distribution) is used to find $\hat{N}$ while the actual distribution of $\hat{\boldsymbol{\Sigma}}_m^{-1/2}\overline{\mathbf{X}}$ is a multivariate t-distribution. Chatterjee (1959, 1960), in fact, uses a multivariate t-distribution in his Stein-type two-stage procedure for accurate multivariate estimation with ellipsoidal accuracy. Chatterjee's procedure gives exact coverage probability; this exactness, however, is achieved only at the cost of considerable computational complexity. Lohr (1988) inflates the covariance estimate by a factor $(1 + l/m)$ to compensate for not knowing $\boldsymbol{\Sigma}$, choosing l so that the double-sampling stopping rule gives coverage probability γ with error $o(m^{-1})$.

The double-sampling stopping rules work very well if the pilot-sample size m has the same order of magnitude as the optimal sample size N. If m is small relative to N, however, the double-sampling stopping time has infinite regret and large variance and thus is inefficient when compared with the purely sequential procedures of Chow and Robbins (1965) and Woodroofe (1977) for one-dimensional accurate estimation and Finster (1986) for multi-dimensional accurate estimation. The triple-sampling procedure of this paper achieves finite regret and second-order asymptotic efficiency by taking two additional samples after the pilot sample rather than just one. As in Hall (1981), we allow for three samples by having the second sample comprise about $100c\%$ $(0 < c < 1)$ of the observations in the second and third samples. N_2, the optimal size of the first and second samples if $\boldsymbol{\Sigma}$ were known, is set equal to $c(N - m) + m$. Then N_2 is estimated after the pilot sample by the stopping time

$$t_2 = [\![c(\hat{N} - m)_+]\!] + m,$$

where $[\![x]\!]$ denotes the smallest integer containing x. After the second sample of $t_2 - m$ observations, the covariance matrix is reestimated using $\hat{\boldsymbol{\Sigma}}_{t_2}$, the least squares estimate of $\boldsymbol{\Sigma}$ using all t_2 observations. Then the size of the third and final sample is $t_3(l) - t_2$, where

$$t_3(l) = [\![N\big((1 + l/t_2)\hat{\boldsymbol{\Sigma}}_{t_2}\big)]\!].$$

Here l compensates for the sequential nature of the procedure. With l defined in (2.3), triple sampling achieves the bounded regret of Simons (1968) and attains coverage probability γ with error $o(N^{-1})$, the same order obtained by Finster (1986). With this small order of error, the asymptotic results for triple sampling apply even to moderate values of N.

Note that accurate estimates of linear combinations of the parameters are a by-product of accurate estimates of the parameters if the accuracy set is a ball. Suppose $P(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \in B_q(d)) = \gamma$, where $B_q(d)$ is the l^q-ball of radius d. Then if $q' = (1 - q^{-1})^{-1}$, an application of Hölder's inequality yields

$$P\big(\big|\mathbf{c}^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\big| \le d\|\mathbf{c}\|_{q'}, \forall\, \mathbf{c} \in \mathscr{R}^p\big) \ge P\big(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_q \le d\big) \ge \gamma.$$

The values $q = 2$ and $q = \infty$ give fixed-accuracy analogues of the Scheffé and Tukey procedures for obtaining simultaneous confidence intervals.

The definition of accuracy used in this paper is that given by Finster (1985, 1986). The techniques used to develop the asymptotic properties for triple sampling, however, are quite different from those of Finster's continuously monitoring procedures or the spherical accuracy procedures in Srivastava (1967) and Srivastava and Bhargava (1979). Finster's results depend on the fact that the stopping time of a purely sequential procedure is the first passage time of a function of a process similar to a random walk. The procedures in this paper are closer in spirit to those of Cox (1952) and Hall (1981), using Taylor series expansions and conditional inference. They extend the results of Mukhopadhyay and Al-Mousawi (1986), who considered accurate multivariate estimation for elliptical accuracy sets when the correlation matrix is known.

2. Triple-sampling procedures for accurate estimation. Accurate estimation is most expensive when the standard deviations for the components of the observations are large relative to the accuracy desired, i.e., when $\boldsymbol{\Sigma}^{-1/2}A$

is small. Following Anscombe (1953), asymptotic results are expressed in terms of N increasing to infinity. Note that N increases to infinity either as $\mathbf{\Sigma} \to \infty$ or as the accuracy set decreases to the empty set. We take $\mathbf{\Sigma} \to \infty$ as $N \to \infty$ to mean that

$$A' \equiv (N\mathbf{\Sigma}^{-1})^{1/2} A \tag{2.1}$$

is a constant set as $N \to \infty$. In other words, $\mathbf{\Sigma} \to \infty$ along a ray. This formulation is consistent with the asymptotic results of Stein (1945) and Chow and Robbins (1965), in which d, the half-width of the confidence interval, tends to zero. If A in (1.2) is replaced by dA, then $N \to \infty$ as $d \to 0$.

The following theorem demonstrates that conditionally on a stopping time τ, the estimate $\overline{\mathbf{X}}_\tau$ has the same distribution it would have if τ were a fixed integer rather than a random variable. The proof of the theorem follows the proofs of Lemmas 1 through 4 in Robbins (1959).

THEOREM 1. *Let τ be an integer-valued stopping time which is a function $(\hat{\mathbf{\Sigma}}_{p+1}, \hat{\mathbf{\Sigma}}_{p+2}, \hat{\mathbf{\Sigma}}_{p+3}, \dots)$. Then the conditional distribution of $\overline{\mathbf{X}}_\tau$ given $\tau = n$ is $\mathscr{N}(\boldsymbol{\theta}, \mathbf{\Sigma}/n)$.*

Theorem 1 implies that conditionally on the stopping time $t_3(l)$, the parameter estimates $\overline{\mathbf{X}}_{t_3(l)}$ are normally distributed with mean $\boldsymbol{\theta}$ and covariance $\mathbf{\Sigma}/t_3(l)$ and hence are unbiased. We now state the main result about second-order properties of the triple-sampling procedure. Throughout, let Φ represent the standard multivariate normal probability measure and adopt the notation

$$\mathbf{S}_x = \operatorname{tr}(\mathbf{S}) - \mathbf{x}^T\mathbf{S}\mathbf{x}$$

and

$$E_R[h(\mathbf{x})] = \int_R h(\mathbf{x})\, d\Phi(\mathbf{x}).$$

Define

$$\mathbf{M} \equiv E_{A'}[\mathbf{I} - \mathbf{x}\mathbf{x}^T]/E_{A'}[\mathbf{I}_x],$$

where $\mathbf{I}$ is the $p \times p$ identity matrix. Note that $\operatorname{tr}\mathbf{M} = 1$ and by Theorem 9.1.25 of Graybill (1983),

$$p^{-1} \le \operatorname{tr}(\mathbf{M}^2) \le 1. \tag{2.2}$$

THEOREM 2. *Let $\mathbf{X}_1, \mathbf{X}_2, \dots$ be independent and identically distributed $\mathscr{N}(\boldsymbol{\theta}, \mathbf{\Sigma})$ random vectors. Let $N = N(\mathbf{\Sigma})$, $t_2 = [\![c(\hat{N} - m)]\!]_+ + m$ and $t_3(l) = [\![N((1 + l/t_2)\hat{\mathbf{\Sigma}}_{t_2})]\!]$, where l is a known constant and the function N is defined in* (1.1). *Assume $m \to \infty$ as a fractional power of N, so that $N = O(m^h)$ for some $h > 1$ but $m/N \to 0$. Let A' be as defined in* (2.1). *Then, as $N \to \infty$,*

(a) *$t_3(l)/N \to 1$ almost surely.*
(b) *For any $q \in \mathscr{R}$, $E([t_3(l)/N]^q) \to 1$.*

(c) $E[t_3(l)] = N + l/c - 2\operatorname{tr}\mathbf{M}^2/c + \frac{1}{2} + [2cE_{A'}[\mathbf{I}_x]]^{-1}E_{A'}$
$\times[2p(\operatorname{tr}\mathbf{M}^2) - 4 + 2p + \mathbf{I}_x[\mathbf{x}^T\mathbf{x}(\operatorname{tr}\mathbf{M}^2 + 1) - 2\mathbf{x}^T\mathbf{M}\mathbf{x}]]$
$+(2c)^{-1}[(4-p)\operatorname{tr}(\mathbf{M}^2) - p - 2] + o(1).$

(d) $E[(t_3(l) - N)^2] = 2N\operatorname{tr}(\mathbf{M}^2)/c + o(N).$

(e) $\sqrt{c}\,(t_3(l) - N)/\sqrt{N}$ *converges to a* $\mathscr{N}(0, 2\operatorname{tr}(\mathbf{M}^2))$ *distribution.*

(f) $P(\overline{\mathbf{X}}_{t_3(l)} - \boldsymbol{\theta} \in A)$

$$= \gamma + (4cN)^{-1}\{[2l - p + c - 2]E_{A'}[\mathbf{I}_x] + E_{A'}(-4 + 2p + \mathbf{I}_x[\mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{M}\mathbf{x}])\} + o(N^{-1}).$$

To attain asymptotically correct coverage probability up to $o(N^{-1})$ terms, we find the value l_3 which solves $P(\overline{\mathbf{X}}_{t_3(l)} - \boldsymbol{\theta} \in A) = \gamma + o(N^{-1})$. Set

$$(2.3)\quad l_3 \equiv -E_{A'}[2p - 4 + \mathbf{I}_x[\mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{M}\mathbf{x}]]/2E_{A'}[\mathbf{I}_x] + p/2 + 1 - c/2.$$

The first term in (2.3) depends upon N and Σ through the set A'. We substitute the estimates $t_3(0)$ and $\hat{\Sigma}_{t_2}$ for N and Σ in (2.3) and define

$$\hat{A}' \equiv \left(t_3(0)\hat{\Sigma}_{t_2}^{-1}\right)^{1/2}A,$$
$$\hat{\mathbf{M}} \equiv E_{\hat{A}'}[\mathbf{I} - \mathbf{x}\mathbf{x}^T]/E_{\hat{A}'}[\mathbf{I}_x]$$

and

$$(2.4)\quad \hat{l}_3 \equiv -E_{\hat{A}'}[2p - 4 + \mathbf{I}_x[\mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\hat{\mathbf{M}}\mathbf{x}]]/2E_{\hat{A}'}[\mathbf{I}_x] + p/2 + 1 - c/2.$$

COROLLARY 1. *Let* $t_3(\hat{l}_3) = [\![N((1 + \hat{l}_3/t_2)\hat{\Sigma}_{t_2})]\!]$, *where* $\hat{l}_3$ *is defined in* (2.4). *Then the results of Theorem* 2 *hold when* $\hat{l}_3$ *is substituted for* l. *In particular, as* $N \to \infty$,

$$\text{(c)}\quad E[t_3(\hat{l}_3)] = N + (\operatorname{tr}\mathbf{M}^2/2c) \times(-p + E_{A'}[2p + \mathbf{I}_x\mathbf{x}^T\mathbf{x}]/E_{A'}[\mathbf{I}_x]) + o(1)$$

and

$$\text{(f)}\quad P(\overline{\mathbf{X}}_{t_3(\hat{l}_3)} - \boldsymbol{\theta} \in A) = \gamma + o(N^{-1}).$$

Theorem 2 and Corollary 1 are proven in Section 3.

Theorem 2 demonstrates that the triple-sampling procedure attains first-order asymptotic efficiency if the pilot-sample size tends to infinity as some fractional power of N, the "best" fixed-sample size. The first-order properties do not depend on the covariance and do not require a correction factor. We may use the estimate $\hat{\Sigma}$ in place of the unknown covariance Σ and still have an asymptotically correct procedure up to first-order asymptotic terms. If γ is 0.95 or 0.99, though, an error of order $o(1)$ can make a substantial difference in the coverage probability unless N is very large indeed. Lavenberg and Sauer (1977) found that sequential stopping rules with only first-order asymptotic consistency perform poorly for relatively small sample sizes. The second-order asymptotic results apply to more moderate values of N. The effects of substituting $\hat{\Sigma}$ for Σ appear in the second-order asymptotic results, particularly in the terms of order $O(N^{-1})$ in the expression for the coverage probability in Theorem 2(f).

The term $-2\,\mathrm{tr}(\mathbf{M}^2)/c$ in the expression for the average sample number shows the effect of optional stopping and appears because $\hat{\Sigma}_{t_2}$ has bias $-2\Sigma^{1/2}\mathbf{M}\Sigma^{1/2}/N_2$. This bias is proven in Lemma 7 of Section 3 and may be heuristically explained as follows. If $\hat{\Sigma}_m$ significantly "overestimates" Σ, then t_2 will overestimate N_2 and the second sample will be large, tending to correct the original estimate of the covariance. Alternatively, if $\hat{\Sigma}_m$ "underestimates" Σ, then t_2 will underestimate N_2. The second sample will thus not contain as many observations to compensate for the bias arising in the first sample, so $\hat{\Sigma}_{t_2}$ will be more likely to err in the same direction as $\hat{\Sigma}_m$. The argument that $\hat{\Sigma}_{t_2}$ is biased also applies to $\hat{\Sigma}_{t_3(l)}$. If one ignores the fact that these quantities are obtained sequentially, substituting $\hat{\Sigma}_{t_3(l)}$ for a fixed-sample estimate of the covariance in, say, an F-test for the significance of one of the means, one would thus obtain more false positive results in repeated sampling than the nominal significance level indicated.

The factor $\mathrm{tr}(\mathbf{M}^2)$ appears in the expression for the variance of the stopping time. The matrix $\mathbf{M}$ shows the effect of the shape and orientation of the standardized accuracy set A′ on the stopping times. If A is spherical and the components of $\mathbf{X}$ are independent, then $\mathbf{M}$ is a diagonal matrix and $2\,\mathrm{tr}(\mathbf{M}^2) = 2/p$. Alternatively, suppose that the components of $\mathbf{X}$ are highly positively correlated. Then most of the variance is accounted for in the first principal component and the stopping times will be essentially determined by the variance of the first principal component. In this case, then, $2\,\mathrm{tr}(\mathbf{M}^2)$ will be close to two, resulting in the variance for the one-dimensional procedures of Cox (1952) and Hall (1981).

The one-dimensional results of Hall follow as special cases of the results in Theorem 2. Let z be the $(1-\gamma)/2$ critical point of the standard normal distribution. Then $A' = [-z, z]$ and $\mathbf{M} = 1$. Evaluating the integrals in Theorem 2, $E[t_3(l)] = N + (l-2)/c + \frac{1}{2} + o(1)$ and $P(\overline{X}_{[\![t_3(l)]\!]} - \theta \in A) = \gamma(2cN\sqrt{2\pi})^{-1}z\exp(-z^2/2)[2l + c - 5 - z^2]$. Thus the triple-sampling procedure which uses $l_3 = (5 + z^2 - c)/2$ will have coverage probability $\gamma + o(N^{-1})$, as obtained by Hall.

Hall recommends using $\frac{1}{2}$ for c. An alternative choice uses the distribution of $\hat{N}$. Since the distribution of $\hat{N}$ is approximately $\mathscr{N}(N, 2N^2\,\mathrm{tr}(\mathbf{M}^2)/m)$, (2.2) implies that $[1 - z_\alpha(2/m)^{1/2}]\hat{N}$ is a conservative $(1-\alpha)$ lower confidence bound for N, where z_α is the appropriate normal percentile. This suggests taking c to be $1 - z_\alpha(2/m)^{1/2}$.

Table 1 compares the properties of multivariate double- and triple-sampling procedures and Finster's (1986) purely sequential procedure. The quantity l_3 is messy to calculate exactly but may be bounded by $p + 2 + pK$, where K is the radius of the smallest sphere which will circumscribe the standardized

TABLE 1
Properties of the double-sampling, triple-sampling and purely sequential procedures for multivariate estimation

	Double sampling	**Triple sampling**	**Purely sequential**
Correction factor l	$l_3 - 2\,\mathrm{tr}(\mathbf{M}^2)$	l_3	$l_3 - \rho$
Regret	rN/m	r/c	r
Asymptotic variance of stopping time	$2(\mathrm{tr}\,\mathbf{M}^2)N^2/m$	$2(\mathrm{tr}\,\mathbf{M}^2)N/c$	$2(\mathrm{tr}\,\mathbf{M}^2)N$
Approximate distribution of stopping time	Linear combination of χ_1^2	Normal	Normal
Coverage probability	$\gamma + o(m^{-1})$	$\gamma + o(N^{-1})$	$\gamma + o(N^{-1})$

The size of the first sample is m, c is the fraction of observations taken in the second sample, N is the "best" fixed-sample stopping rule, ρ corrects for the discreteness of the purely sequential stopping rule and $r = (\mathrm{tr}\,\mathbf{M}^2/2)\{-p + [E_{A'}[\mathbf{I}_x]]^{-1}E_{A'}[2p + \mathbf{I}_x\mathbf{x}^T\mathbf{x}]\}$.

accuracy set A'. In practice, we may use the radius of the smallest sphere circumscribing the sample standardized accuracy set $[t_3(0)\hat{\boldsymbol{\Sigma}}_{t_2}^{-1}]^{1/2}A$ instead of K.

3. Proofs. Throughout, assume without loss of generality that $\boldsymbol{\theta} = \mathbf{0}$ and that $\boldsymbol{\Sigma} = \mathbf{I}$. Theorem 1 implies that $P(\overline{\mathbf{X}}_{t_3(l)} \in A) = E[f(t_3(l), \mathbf{I})]$, where f is defined in (1.2). The coverage probability may then be evaluated using the moments of $t_3(l)$ by taking a Taylor series expansion about the first argument,

$$P\big(\overline{\mathbf{X}}_{t_3(l)} \in A\big) = f(N, \mathbf{I}) + f_1(N, \mathbf{I})E[t_3(l) - N] + E\Big[f_{11}(n^*, \mathbf{I})\big(t_3(l) - N\big)^2\Big]/2. \tag{3.1}$$

Here n^* is between $t_3(l)$ and N, and f_1 and f_{11} denote the first and second partial derivatives of f with respect to the first argument.

Recall that $t_3(l) = [\![N(\mathbf{W})]\!]$, with $\mathbf{W} = (1 + l/t_2)\hat{\boldsymbol{\Sigma}}_{t_2}$ and $N = N(\mathbf{I})$. We find the moments of $t_3(l)$ via a Taylor series expansion of $N((1 + l/t_3)\hat{\boldsymbol{\Sigma}}_{t_3})$ about $\mathbf{I}$, using Fréchet derivatives to ensure that all matrices will be positive definite. Define the function

$$\varkappa(\varepsilon) = N(\varepsilon\mathbf{W} + (1 - \varepsilon)\mathbf{I}), \tag{3.2}$$

for positive definite $\mathbf{W}$ and $0 \le \varepsilon \le 1$. Here $[[\varkappa(1)]] = t_3(l)$ and $\varkappa(0) = N$, so the moments of $t_3(l)$ may be approximated using a Taylor series expansion of $\varkappa$ about 0. The derivatives of $\varkappa(\varepsilon)$ are more easily evaluated and bounded in a different coordinate system. Let $\boldsymbol{\Lambda}$ be the matrix of eigenvalues of $\mathbf{W}$ and $\mathbf{P}$ the matrix of eigenvectors of $\mathbf{W}$. Then $\varepsilon\mathbf{W} + (1 - \varepsilon)\mathbf{I} = \mathbf{P}^T\mathbf{L}(\varepsilon)\mathbf{P}$, where

$$\mathbf{L}(\varepsilon) = \varepsilon[\boldsymbol{\Lambda} - \mathbf{I}] + \mathbf{I}. \tag{3.3}$$

Define the set

$$A^*(n,\varepsilon) \equiv \left[n\mathbf{L}^{-1}(\varepsilon)\right]^{1/2}\mathbf{P}^T A \tag{3.4}$$

[throughout, A^* without arguments refers to $A^*(\varkappa(\varepsilon),\varepsilon)$] and the function

$$g(n,\varepsilon) \equiv f(n, \varepsilon\mathbf{W} + (1-\varepsilon)\mathbf{I}), \tag{3.5}$$

Then $g(n,\varepsilon)$ may be rewritten as

$$\begin{aligned} g(n,\varepsilon) &= \int_{\mathbf{P}^T A} (n/2\pi)^{p/2} |\mathbf{L}(\varepsilon)|^{-1/2} \exp\{-(n/2)\mathbf{x}^T\mathbf{L}^{-1}(\varepsilon)\mathbf{x}\}\,d\mathbf{x} \\ &= \Phi[A^*(n,\varepsilon)] \end{aligned}$$

and the derivatives of g and $\varkappa$ found directly.

Lemma 1. *Let $\varkappa(\varepsilon)$ and $g(n,\varepsilon)$ be defined in* (3.2) *through* (3.5) *and let* $\mathbf{K}$ *and* $\mathbf{H}$ *denote*

$$\mathbf{K}(\varepsilon) \equiv \mathbf{L}^{-1/2}(\varepsilon)[\mathbf{\Lambda} - \mathbf{I}]\mathbf{L}^{-1/2}(\varepsilon) \tag{3.6}$$

and

$$\mathbf{H}(\varepsilon) \equiv (\varkappa'(\varepsilon)/\varkappa(\varepsilon))\mathbf{I} - \mathbf{K}(\varepsilon), \tag{3.7}$$

respectively. Then

$$\begin{aligned} g_1(n,\varepsilon) &= E_{A^*(n,\varepsilon)}[\mathbf{I}_x]/2n, \\ g_2(n,\varepsilon) &= -E_{A^*(n,\varepsilon)}[\mathbf{K}_x]/2, \\ g_{11}(n,\varepsilon) &= E_{A^*(n,\varepsilon)}\left[\mathbf{I}_x^2 - 2p\right]/4n^2, \\ g_{12}(n,\varepsilon) &= -E_{A^*(n,\varepsilon)}\left[\mathbf{K}_x\mathbf{I}_x - 2\mathbf{x}^T\mathbf{K}\mathbf{x}\right]/4n, \\ g_{22}(n,\varepsilon) &= E_{A^*(n,\varepsilon)}\left[\mathbf{K}_x^2 + 2\,\mathrm{tr}\,\mathbf{K}^2 - 4\mathbf{x}^T\mathbf{K}^2\mathbf{x}\right]/4, \\ g_{111}(n,\varepsilon) &= E_{A^*(n,\varepsilon)}\left[\mathbf{I}_x^3 - 6p\mathbf{I}_x + 8p\right]/8n^3, \end{aligned}$$

$$\varkappa'(\varepsilon) = \varkappa(\varepsilon)E_{A^*}[\mathbf{K}_x]/E_{A^*}[\mathbf{I}_x],$$

$$\varkappa''(\varepsilon) = \varkappa(\varepsilon)\{2\gamma\,\mathrm{tr}(\mathbf{H}^2) - E_{A^*}[\mathbf{H}_x^2 - 4(\mathbf{HK})_x]\}/2E_{A^*}[\mathbf{I}_x],$$

$$\begin{aligned} \varkappa'''(\varepsilon) &= \varkappa(\varepsilon)[4E_{A^*}[\mathbf{I}_x]]^{-1}E_{A^*}\big[\mathrm{tr}[24\{\varkappa''(\varepsilon)/\varkappa(\varepsilon)\}\mathbf{H} - 8\mathbf{H}^3 - 24\mathbf{H}^2\mathbf{K}] \\ &\quad + 12\mathbf{H}_x(\mathbf{HK} - \{\varkappa''(\varepsilon)/\varkappa(\varepsilon)\}\mathbf{I})_x + 24(\{\varkappa''(\varepsilon)/\varkappa(\varepsilon)\}\mathbf{K} - \mathbf{HK}^2)_x - \mathbf{H}_x^3\big]. \end{aligned}$$

The integral $E_{A^*}[\mathbf{I}_x]$, appearing in the denominators of the expressions for $\varkappa'(\varepsilon)$, $\varkappa''(\varepsilon)$ and $\varkappa'''(\varepsilon)$, is always greater then zero but can be small. To facilitate finding bounds for the derivatives of $\varkappa$, we apply Stokes' theorem, as quoted in Spivak (1965), to express them as integrals over the boundary of A^*. Here, $d\mathbf{x}^{(i)}$ is written for $d\mathbf{x}_1 d\mathbf{x}_2 \cdots d\mathbf{x}_{i-1} d\mathbf{x}_{i+1} \cdots d\mathbf{x}_p$, ∂A^* represents the boundary of A^* and $d\Phi^{(i)}(\mathbf{x}) = (2\pi)^{-p/2}\exp(-\mathbf{x}^T\mathbf{x}/2)\,d\mathbf{x}^{(i)}$.

Lemma 2. *Let $\varkappa(\varepsilon)$, $g(n,\varepsilon)$, $\mathbf{K}(\varepsilon)$ and $\mathbf{H}(\varepsilon)$ be defined in* (3.2) *through* (3.7), *let g_1 denote $g_1(\varkappa(\varepsilon),\varepsilon)$, and let $\mathbf{H}_i$ and $\mathbf{K}_i$ denote the ith diagonal* *entries of* **H** *and* **K**. *Then*

(a) $g_1 = [2\varkappa(\varepsilon)]^{-1}\Sigma(-1)^{i-1}\int_{\partial A^*} x_i \, d\Phi^{(i)}(\mathbf{x})$.

(b) $$\varkappa''(\varepsilon) = [4g_1]^{-1}\Big\{\sum (-1)^{i-1}\mathbf{H}_i \int_{\partial A^*} x_i\big[\mathbf{x}^T\mathbf{H}\mathbf{x} - \operatorname{tr}(\mathbf{H}) + 2\mathbf{H}_i + 4\mathbf{K}_i\big]\, d\Phi^{(i)}(\mathbf{x})\Big\}.$$

(c) $$\begin{aligned}\varkappa'''(\varepsilon) = [8g_1]^{-1} &\times\Big\{-12\sum(-1)^{i-1}\mathbf{H}_i\int_{\partial A^*} x_i\big[\mathbf{x}^T\mathbf{H}\mathbf{K}\mathbf{x} + 2\mathbf{H}_i\mathbf{K}_i\big]\, d\Phi^{(i)}(\mathbf{x})\\ &+12\sum(-1)^{i-1}\mathbf{H}_i\{\varkappa''(\varepsilon)/\varkappa(\varepsilon)\}\int_{\partial A^*} x_i[\mathbf{x}^T\mathbf{x} + 2]\, d\Phi^{(i)}(\mathbf{x})\\ &+24\sum(-1)^{i-1}\mathbf{K}_i(\{\varkappa''(\varepsilon)/\varkappa(\varepsilon)\} - \mathbf{H}_i\mathbf{K}_i)\int_{\partial A^*} x_i\, d\Phi^{(i)}(\mathbf{x})\\ &-\sum(-1)^{i-1}\mathbf{H}_i\int_{\partial A^*} x_i\Big[(\mathbf{x}^T\mathbf{H}\mathbf{x})^2 + (4\mathbf{H}_i - 2\operatorname{tr}[\mathbf{H}])\mathbf{x}^T\mathbf{H}\mathbf{x}\\ &+8\mathbf{H}_i^2 + 4\operatorname{tr}[\mathbf{H}^2] - 4\mathbf{H}_i\operatorname{tr}[\mathbf{H}]\Big]\, d\Phi^{(i)}(\mathbf{x})\Big\}.\end{aligned}$$

We are now in a position to bound the derivatives of $\varkappa(\varepsilon)$ for all values of ε between 0 and 1. Let

$$\tilde{\lambda} \equiv \max(\lambda_1, 1) \tag{3.8}$$

and

$$\underline{\lambda} \equiv \min(\lambda_p, 1), \tag{3.9}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ are the eigenvalues of **W**. For any matrix **C**, let $\|\mathbf{C}\|_\infty$ represent the supremum norm of a matrix; $\|\mathbf{C}\|_\infty \equiv \max|\mathbf{C}_{ij}|$.

LEMMA 3. *Let* $0 \leq \varepsilon \leq 1$ *and let* $\mathbf{L}(\varepsilon)$, $\mathbf{K}(\varepsilon)$, $\mathbf{H}(\varepsilon)$, $\tilde{\lambda}$ *and* $\underline{\lambda}$ *be as defined in* (3.3) *and* (3.6) *through* (3.9). *Then*

(a) $\underline{\lambda}\mathbf{I} \leq \mathbf{L}(\varepsilon) \leq \tilde{\lambda}\mathbf{I}$.

(b) $\underline{\lambda}N \leq \varkappa(\varepsilon) \leq \tilde{\lambda}N$.

(c) $|\varkappa'(\varepsilon)/\varkappa(\varepsilon)| \leq \|\mathbf{K}(\varepsilon)\|_\infty$.

(d) $\|\mathbf{H}(\varepsilon)\| \leq 2\|\mathbf{K}(\varepsilon)\|_\infty$.

(e) *There exists a constant* K, *independent of* ε, *such that* $a \in A^*(N, 0)$ *implies* $a^Ta \leq K$ *and* $a \in A^*(\varkappa(\varepsilon), \varepsilon)$ *implies* $a^Ta \leq (\tilde{\lambda}/\underline{\lambda})K$.

(f) $|\varkappa''(\varepsilon)| \leq \varkappa(\varepsilon)(\tilde{\lambda}/\underline{\lambda})K_1\|\mathbf{K}(\varepsilon)\|_\infty^2$, *where* $K_1 = 2p + 8 + 2K$.

(g) $|\varkappa'''(\varepsilon)| \leq \varkappa(\varepsilon)\|\mathbf{K}(\varepsilon)\|_\infty^3(\tilde{\lambda}/\underline{\lambda})^2K_2$, *where* K_2 *is a constant*.

(h) $\|\mathbf{K}(\varepsilon)\|_\infty \leq \underline{\lambda}^{-1}\|\mathbf{W} - \mathbf{I}\|$ *for any matrix norm* $\|\cdot\|$.

PROOF. (a) Follows immediately from (3.3).

(b) Suppose $\varkappa(\varepsilon) > \tilde{\lambda}N$. Then using (a),

$$\gamma = \Phi\Big[\big[\varkappa(\varepsilon)\mathbf{L}^{-1}(\varepsilon)\big]^{1/2}\mathbf{P}^TA\Big] \geq \Phi\Big[\big[\mathbf{I}\varkappa(\varepsilon)/\tilde{\lambda}\big]^{1/2}\mathbf{P}^TA\Big] > \Phi[N^{1/2}\mathbf{P}^TA] = \gamma,$$

a contradiction. Therefore $\varkappa(\varepsilon) \leq \tilde{\lambda}N$; the other inequality is proven similarly.

(c) From Lemma 1,

$$\varkappa'(\varepsilon) = \varkappa(\varepsilon)\left[\sum \mathbf{K}_i(\varepsilon) E_{A^*}[1 - x_i^2]\right]\left[\sum E_{A^*}[1 - x_i^2]\right]^{-1} \le \varkappa(\varepsilon)\|K(\varepsilon)\|_\infty.$$

The integrals $E_{A^*}[1 - x_i^2]$ are all strictly positive since A^* is an accuracy set.

(d) Follows immediately from part (c), the definition of $\mathbf{H}$ and the triangle inequality.

(e) Using (a), (c) and the definition of $A' = N^{1/2}A$,

$$A^* = [\varkappa(\varepsilon)/N]^{1/2}\mathbf{L}^{-1/2}(\varepsilon)\mathbf{P}^T A' \subset (\tilde{\lambda}/\underline{\lambda})^{1/2}\mathbf{P}^T A'.$$

Recall that the standardized accuracy set A' is a constant, bounded set as N tends to infinity and is thus contained in a ball of radius $\sqrt{K}$ for some $K < \infty$. Since $\mathbf{P}$ is orthogonal, $a^T a \le K$ if $a \in A^*(N, 0)$, and $a^T a \le (\tilde{\lambda}/\underline{\lambda})K$ if $a \in A^*(\varkappa(\varepsilon), \varepsilon)$ and $0 \le \varepsilon \le 1$.

(f) From Lemma 2,

$$\varkappa''(\varepsilon) = \varkappa(\varepsilon)\left\{2\sum (-1)^{i-1}\int_{\partial A^*} x_i \, d\Phi^{(i)}(\mathbf{x})\right\}^{-1}$$

$$\times\left\{\sum (-1)^{i-1}\mathbf{H}_i\int_{\partial A^*} x_i\left[-\mathrm{tr}(\mathbf{H}) + 2\mathbf{H}_i + 4\mathbf{K}_i + \mathbf{x}^T\mathbf{H}\mathbf{x}\right] d\Phi^{(i)}(\mathbf{x})\right\}.$$

Now by (d),

$$|\mathbf{H}_i[-\mathrm{tr}(\mathbf{H}) + 2\mathbf{H}_i + 4\mathbf{K}_i]| \le 4(p+4)\|\mathbf{K}(\varepsilon)\|_\infty^2 \quad \text{for } i = 1, \dots, p.$$

Also, Stokes' theorem implies that

$$(-1)^{i-1}\int_{\partial A^*} x_i x_j^2 \, d\Phi^{(i)}(\mathbf{x}) \ge E_{A^*}\left[x_j^2(1 - x_i^2)\right] > 0,$$

so

$$\left|\sum\sum (-1)^{i-1}\mathbf{H}_i\mathbf{H}_j\int_{\partial A^*} x_i x_j^2 \, d\Phi^{(i)}(\mathbf{x})\right|$$

$$\le \|\mathbf{H}\|_\infty^2 \sum (-1)^{i-1}\int_{\partial A^*} x_i \mathbf{x}^T\mathbf{x} \, d\Phi^{(i)}(\mathbf{x}).$$

Let $\psi(\mathbf{u})$, $\mathbf{u} \in \mathscr{R}^{p-1}$, be an orientation-preserving parameterization of the boundary ∂A^*, with Γ the region of integration for $\mathbf{u}$. Then

$$\sum (-1)^{i-1}\int_{\partial A^*} x_i \mathbf{x}^T\mathbf{x} \, d\Phi^{(i)}(\mathbf{x}) = \int_\Gamma \psi^T\psi \exp(-\psi^T\psi/2)\mathbf{J}(\mathbf{u}) \, d\mathbf{u},$$

where $\mathbf{J}(\mathbf{u})$ is the Jacobian of the transformation. Now $\mathbf{J}(\mathbf{u})$ is always positive; hence for all $\mathbf{u}$ in Γ,

$$\psi^T(\mathbf{u})\psi(\mathbf{u})\exp(-\psi^T(\mathbf{u})\psi(\mathbf{u})/2)\mathbf{J}(\mathbf{u}) \le (\tilde{\lambda}/\underline{\lambda})K \exp(-\psi^T(\mathbf{u})\psi(\mathbf{u})/2)\mathbf{J}(\mathbf{u}),$$

by (e) since $\psi^T(\mathbf{u})\psi(\mathbf{u}) \in A^*$ whenever $\mathbf{u} \in \Gamma$. Thus

$$\left|\sum (-1)^{i-1}\mathbf{H}_i \int_{\partial A^*} x_i \mathbf{x}^T \mathbf{H}\mathbf{x}\, d\Phi^{(i)}(\mathbf{x})\right|$$

$$\leq (\tilde{\lambda}/\underline{\lambda}) K \|\mathbf{H}\|_\infty^2 \sum (-1)^{i-1} \int_{\partial A^*} x_i \, d\Phi^{(i)}(\mathbf{x}).$$

As a result, $|\varkappa''(\varepsilon)/\varkappa(\varepsilon)| \leq (2p + 8 + 2(\tilde{\lambda}/\underline{\lambda})K)\|\mathbf{K}(\varepsilon)\|_\infty^2$.

(g) Proven similarly to (f), using Lemma 2 and parts (d) and (f) of Lemma 3 to show that

$$|\varkappa'''(\varepsilon)| \leq \varkappa(\varepsilon)\|\mathbf{K}(\varepsilon)\|_\infty^3 (\tilde{\lambda}/\underline{\lambda})^2 \left[14K^2 + (104 + 16p)K + 196 + 52p\right].$$

(h) Follows from (3.6), part (a), and Theorem 5.6.7 of Graybill (1983). □

LEMMA 4. *Suppose $n > 0$ and $0 < \varepsilon < 1$. Then*

(a) $0 < g_1(n, \varepsilon) \leq p/(2n)$.
(b) $|g_{11}(n, \varepsilon)| \leq (2n)^{-2}[p^2 + 2p + (n/\varkappa(\varepsilon))^2(\tilde{\lambda}/\underline{\lambda})^2 K^2]$.
(c) $|g_{111}(n, \varepsilon)| \leq (2n)^{-3} p[4p + (n/\varkappa(\varepsilon))(\tilde{\lambda}/\underline{\lambda})K]^2$.

PROOF.

(b)
$$|g_{11}(n, \varepsilon)| = \left|(2n)^{-2} E_{A^*(n,\varepsilon)}[\mathbf{I}_x^2 - 2p]\right|$$
$$\leq (2n)^{-2}\left[p^2 + 2p + E_{(n/\varkappa(\varepsilon))^{1/2}A^*}\left[(\mathbf{x}^T\mathbf{x})^2\right]\right].$$

Now by Lemma 3(e), $\mathbf{x}^T\mathbf{x} \leq (\tilde{\lambda}/\underline{\lambda})K$ on A^*. Hence

$$|g_{11}(n, \varepsilon)| \leq (2n)^{-2}\left[2p + p^2 + (n/\varkappa(\varepsilon))^2 (\tilde{\lambda}/\underline{\lambda})^2 K^2\right].$$

(c) As in the proof of part (b),

$$|g_{111}(n, \varepsilon)| \leq (2n)^{-3} \int_{A^*(n,\varepsilon)} \left[p\left[p^2 + (\mathbf{x}^T\mathbf{x})^2\right] + 6p\left[p + \mathbf{x}^T\mathbf{x}\right] + 8p\right] d\Phi(\mathbf{x})$$
$$\leq (2n)^{-3}\left[15p^3 + p(n/\varkappa(\varepsilon))^2(\tilde{\lambda}/\underline{\lambda})^2 K^2 + 6p(n/\varkappa(\varepsilon))(\tilde{\lambda}/\underline{\lambda})K\right].$$ □

To find the expectations of the derivatives of $\varkappa$ we work in the usual inner product space of $p \times p$ matrices, with

$$\langle \mathbf{D}_1, \mathbf{D}_2 \rangle = \operatorname{tr}[\mathbf{D}_1\mathbf{D}_2] = (\operatorname{vec}\mathbf{D}_1)^T(\operatorname{vec}\mathbf{D}_2).$$

Recall that if $\mathbf{D}$ is any $p \times p$ matrix, then

$$\operatorname{vec}\mathbf{D} = \left[\mathbf{D}_{11}\mathbf{D}_{21} \cdots \mathbf{D}_{p1}\mathbf{D}_{12}\mathbf{D}_{22} \cdots \mathbf{D}_{p2}\mathbf{D}_{13} \cdots \mathbf{D}_{pp}\right]^T.$$

Let $\otimes$ represent the left Kronecker product on matrices and let $\mathbf{C}$ be the $p^2 \times p^2$ commutation matrix. In the following, let $\Rightarrow$ denote convergence in distribution.

LEMMA 5. *Suppose that* $n\,\mathrm{vec}(\mathbf{W}-\mathbf{I}) \Rightarrow \mathscr{N}(\mathbf{0}, \mathbf{I}\otimes\mathbf{I}+\mathbf{C})$ *for some* n *increasing to infinity and that the moments of* $\mathbf{W}$ *are bounded. Then for* $j = 0, 1, \ldots, (m-p)/2$ *and* $i, k = 0, 1, 2, \ldots,$

$$\text{(a)}\qquad E\left[\tilde{\lambda}^i \underline{\lambda}^{-j} n^k \|\mathbf{W}-\mathbf{I}\|_\infty^k\right] \le 2^k p^{j+1}(\alpha_{2k})^{1/2} + o(1),$$

where α_{2k} *is the* $2k$th *moment of the standard normal distribution.*

$$\text{(b)}\qquad E\left[(n\varkappa'(0)/N)^2\right] = 2\,\mathrm{tr}[\mathbf{M}^2] + o(1).$$

$$\begin{aligned}\text{(c)}\quad E\left[n^2\varkappa''(0)/N\right] = {} & (4-p)\mathrm{tr}(\mathbf{M}^2) - p - 2 \\ & + E_{A'}\big[2p(\mathrm{tr}\,\mathbf{M}^2) - 4 + 2p \\ & \qquad + \mathbf{I}_x\left[\mathbf{x}^T\mathbf{x}(\mathrm{tr}\,\mathbf{M}^2 + 1) - 2\mathbf{x}^T\mathbf{M}\mathbf{x}\right]\big] \big/ E_{A'}[\mathbf{I}_x] \\ & + o(1).\end{aligned}$$

PROOF. Let $\mathbf{Q} = \mathbf{W} - \mathbf{I}$ and let $\mathbf{R} = (\varkappa'(0)/N)\mathbf{I} - \mathbf{Q}$.

(a) Note that $\underline{\lambda}^{-1} \le \mathrm{tr}[\mathbf{W}^{-1}]$. Since $n\,\mathrm{vec}(\mathbf{W}-\mathbf{I}) \Rightarrow \mathscr{N}(\mathbf{0}, \mathbf{I}\otimes\mathbf{I}+\mathbf{C})$, $n(\mathrm{tr}[\mathbf{W}^{-1}] - p) \Rightarrow \mathscr{N}(0, 2p)$ by the delta method. Thus by dominated convergence, $E[\underline{\lambda}^{-j}] \le E[(\mathrm{tr}\,\mathbf{W}^{-1})^j] = p^j + o(1)$. The entry $(n\mathbf{Q}_{ij})$ has either $\mathscr{N}(0,1)$ or $\mathscr{N}(0,2)$ as its limiting distribution, so dominated convergence implies that $E[n^{2k}\|\mathbf{Q}\|_\infty^{2k}] \le 2^{2k}p^2\alpha_{2k} + o(1)$. Thus by the Cauchy–Schwarz inequality, (3.8) and (3.9),

$$\begin{aligned}E\left[\tilde{\lambda}^i\underline{\lambda}^{-j}n^k\|\mathbf{Q}\|_\infty^k\right] &\le \left\{E\left[\underline{\lambda}^{-2j}\right]E\left[(1+\|\mathbf{Q}\|_\infty)^{2i}n^{2k}\|\mathbf{Q}\|_\infty^{2k}\right]\right\}^{1/2} \\ &\le \left\{\left[p^{2j}+o(1)\right]\left[2^{2k}p^2\alpha_{2k}+o(1)\right]\right\}^{1/2}.\end{aligned}$$

(b) Changing variables, we rewrite $\varkappa'(0) = N\,\mathrm{tr}(\mathbf{QM})$. The asymptotic variance of $n\,\mathrm{tr}(\mathbf{QM}) = n(\mathrm{vec}\,\mathbf{M})^T(\mathrm{vec}\,\mathbf{Q})$ is

$$(\mathrm{vec}\,\mathbf{M})^T(\mathbf{I}\otimes\mathbf{I}+\mathbf{C})(\mathrm{vec}\,\mathbf{M}) = 2\,\mathrm{tr}[\mathbf{M}^2].$$

(c) Again changing variables and simplifying,

$$n^2\varkappa''(0)/N = n^2\left[2\gamma\,\mathrm{tr}(\mathbf{R}^2) - E_{A'}[\mathbf{R}_x^2]\right]/2E_{A'}[\mathbf{I}_x] + 2n^2\,\mathrm{tr}(\mathbf{RQM}).$$

We find the expectation of each term. The asymptotic variance of $(n\,\mathrm{vec}\,\mathbf{R})$ is $2[(\mathrm{tr}\,\mathbf{M}^2)(\mathrm{vec}\,\mathbf{I})(\mathrm{vec}\,\mathbf{I})^T - (\mathrm{vec}\,\mathbf{M})(\mathrm{vec}\,\mathbf{I})^T - (\mathrm{vec}\,\mathbf{I})(\mathrm{vec}\,\mathbf{M})^T] + \mathbf{I}\otimes\mathbf{I} + \mathbf{C}$. Thus, by dominated convergence,

$$\begin{aligned}E\left[n^2\,\mathrm{tr}(\mathbf{R}^2)\right] &= \mathrm{tr}\,E\left[n^2(\mathrm{vec}\,\mathbf{R})(\mathrm{vec}\,\mathbf{R})^T\right] \\ &= 2p(\mathrm{tr}\,\mathbf{M}^2) - 4 + p(p+1) + o(1).\end{aligned}$$

Also,

$$E\{n^2E_{A'}[\mathbf{R}_x^2]\} = 2(\mathrm{tr}\,\mathbf{M}^2)E_{A'}[\mathbf{I}_x^2] - 4E_{A'}[\mathbf{I}_x\mathbf{M}_x] + 2E_{A'}\left[p - 2\mathbf{x}^T\mathbf{x} + (\mathbf{x}^T\mathbf{x})^2\right].$$

Finally, part (b) implies that

$$\begin{aligned}E\left[n^2\,\mathrm{tr}(\mathbf{RQM})\right] &= E\left[\{n\,\mathrm{tr}(\mathbf{MQ})\}^2 - n^2\,\mathrm{tr}(\mathbf{MQ}^2)\right] \\ &= 2\,\mathrm{tr}(\mathbf{M}^2) - (p+1) + o(1). \qquad \square\end{aligned}$$

The moments of $(1 + l/t_2)\hat{\boldsymbol{\Sigma}}_{t_2}$ are needed to evaluate $E[\kappa(0)]$. Because of the sequential nature of the procedure, $\hat{\boldsymbol{\Sigma}}_{t_2}$ does not follow a Wishart distribution, but in fact is a systematically biased estimate of $\boldsymbol{\Sigma}$.

LEMMA 6. *Let $\overline{\mathbf{X}}_{(1)} = \overline{\mathbf{X}}_m$, $\hat{\boldsymbol{\Sigma}}_{(1)} = \hat{\boldsymbol{\Sigma}}_m$ and let $\overline{\mathbf{X}}_{(2)}$ and $\hat{\boldsymbol{\Sigma}}_{(2)}$ be the least squares estimates of the mean and covariance using only the observations in the second sample. Then*

(a) $$\hat{\boldsymbol{\Sigma}}_{t_2} = (t_2 - 1)^{-1}\Big\{(m-1)\hat{\boldsymbol{\Sigma}}_{(1)} + (t_2 - m - 1)\hat{\boldsymbol{\Sigma}}_{(2)} + m(t_2 - m)t_2^{-1}\big[\overline{\mathbf{X}}_{(1)} - \overline{\mathbf{X}}_{(2)}\big]\big[\overline{\mathbf{X}}_{(1)} - \overline{\mathbf{X}}_{(2)}\big]^T\Big\}.$$

(b) $$E\big[(t_2 - 1)\hat{\boldsymbol{\Sigma}}_{t_2}\big|\hat{\boldsymbol{\Sigma}}_m\big] = (m-1)(\hat{\boldsymbol{\Sigma}}_m - \mathbf{I}) + (t_2 - 1)\mathbf{I}.$$

LEMMA 7. *Suppose the conditions of Theorem 2 hold. Then*

(a) $$E\big[|t_2 - N_2|^j t_2^{-k}\big] = o(N^{j-k}) \quad \text{for } k = 0, 1, 2, \dots, [\![(m-p)/2]\!] - 1,\ j \geq 1.$$

(b) $E\big[t_2^{-k}\big] = N_2^{-k} + o(N^{-k})$ *for* $k = 1, 2, \dots, [\![(m-p)/2]\!] - 1$.

(c) $$E\big[\hat{\boldsymbol{\Sigma}}_{t_2}\big] = \mathbf{I} - 2\mathbf{M}/N_2 + o(N^{-1}).$$

PROOF. Let

$$D \equiv \{\hat{N} > m\}$$

and let

$$C \equiv \big\{\|\hat{\boldsymbol{\Sigma}}_m - \mathbf{I}\|_\infty \leq \tfrac{1}{2}\big\}.$$

Lemma 3(b), with $\mathbf{W} = \hat{\boldsymbol{\Sigma}}_m$, implies that $|\hat{N} - N| I_C \leq N/2$. Using Cramér's theorem on large deviations [see Varadhan (1984)],

$$P\{C^c\} \leq 2p^2 \exp\{-(m-1)/24\}.$$

Hence

$$P\{D^c\} \leq P\big\{\{|\hat{N} - N| > N - m\} \cap C\big\} + P\{C^c\} \leq 2p^2 \exp\{-(m-1)/24\},$$

for sufficiently large N. Since $m \to \infty$ as a fractional power of N, we have

$$E\big[|t_2 - N_2|^j / t_2^k I_{D^c}\big] = (c(N-m)/m)^j P\{D^c\} = o(N^{j-k})$$

and

$$E\big[|t_2 - N_2|^j t_2^{-k} I_D\big] = E\Big[\big|[\![c(\hat{N} - m)]\!] - c(N-m)\big|^j / \{[\![c(\hat{N} - m)]\!] + m\}^k I_D\Big] \leq E\Big[(1 + c|\hat{N} - N|)^j / (c\hat{N})^k\Big].$$

$E[|\hat{N} - N|^j / \hat{N}^k] = o(N^{j-k})$ by Theorem 2 of Lohr (1988), completing the proof of (a).

Part (b) follows from (a) by applying the binomial theorem to $E[t_2^{-k}] = N_2^{-k}E[\{1 - (t_2 - N_2)/t_2\}^k]$.

For part (c), Lemma 6 and successive conditioning imply that

$$E\big[\hat{\boldsymbol{\Sigma}}_{t_2}\big] = \mathbf{I} + E\big[(t_2 - 1)^{-1}(m - 1)\big(\hat{\boldsymbol{\Sigma}}_m - \mathbf{I}\big)\big]$$
$$= \mathbf{I} + (m - 1)N_2^{-2}E\big[\big(\hat{\boldsymbol{\Sigma}}_m - \mathbf{I}\big)\big\{-(t_2 - N_2) + (t_2 - 1)^{-1}(t_2 - N_2 - 1)^2\big\}\big].$$

Now

$$E\big[\big(\hat{\boldsymbol{\Sigma}}_m - \mathbf{I}\big)(t_2 - N_2)\big] = cE\big[\big(\hat{\boldsymbol{\Sigma}}_m - \mathbf{I}\big)(\hat{N} - N)\big] + o(1).$$

Write

$$(\hat{N} - N) = n'(0) + n''(\varepsilon^*)/2 = \operatorname{tr}\big[N\mathbf{M}\big(\hat{\boldsymbol{\Sigma}}_m - \mathbf{I}\big)\big] + n''(\varepsilon^*)/2,$$

where $n(\varepsilon)$ is defined in (3.2) with $\mathbf{W} = \hat{\boldsymbol{\Sigma}}_m$ and where ε^* is between 0 and 1. Then $E[(\hat{\boldsymbol{\Sigma}}_m - \mathbf{I})(t_2 - N_2)] = 2cN(m - 1)^{-1}\mathbf{M} + r$, with $|r| \le E[N\tilde{\lambda}^2\underline{\lambda}^{-3} \|\hat{\boldsymbol{\Sigma}}_m - \mathbf{I}\|_\infty^3] = o(N/m)$ by Lemmas 3 and 5(a).

To show that $E[(\hat{\boldsymbol{\Sigma}}_m - \mathbf{I})(t_2 - 1)^{-1}(t_2 - N_2 - 1)^2] = o(N/m)$, let $\delta = m^{-1/2}$ and let $B \equiv \{\|\hat{\boldsymbol{\Sigma}}_m - \mathbf{I}\|_\infty \le \delta\} \cap D$. By Cramér's Theorem, $P\{B^c\} \le (2p + p^2)\exp\{-m^{1/4}/12\}$ and $E[\|\hat{\boldsymbol{\Sigma}}_m - \mathbf{I}\|_\infty(t_2 - 1)^{-1}(t_2 - N_2 - 1)^2 I_B] \le \delta^3 N_2/(1 - \delta)$, so $E[\|\hat{\boldsymbol{\Sigma}}_m - \mathbf{I}\|_\infty(t_2 - 1)^{-1}(t_2 - N_2 - 1)^2] = o(N/m)$. □

PROOF OF THEOREM 2. Recall from (3.2) that for triple sampling,

$$n(\varepsilon) = N\Big(\varepsilon(1 + l/t_2)\hat{\boldsymbol{\Sigma}}_{t_2} + (1 - \varepsilon)\mathbf{I}\Big), \tag{3.10}$$

so that Lemmas 1 through 5 may be applied with $\mathbf{W} = (1 + l/t_2)\hat{\boldsymbol{\Sigma}}_{t_2}$. For any constant n, Graybill [(1983), Theorem 10.10.1] implies that $E[\hat{\boldsymbol{\Sigma}}_n] = \mathbf{I}$ and $\operatorname{Cov}(\operatorname{vec}\hat{\boldsymbol{\Sigma}}_n) = (n - 1)^{-1}(\mathbf{I} \otimes \mathbf{I} + \mathbf{C})$. Thus Anscombe's (1952) theorem implies that

$$\sqrt{N_2}\,\operatorname{vec}\big[(1 + l/t_2)\hat{\boldsymbol{\Sigma}}_{t_2} - \mathbf{I}\big] \Rightarrow \mathscr{N}(\mathbf{0}, \mathbf{I} \otimes \mathbf{I} + \mathbf{C}). \tag{3.11}$$

Dominated convergence may be applied throughout the proof; the moments of $\hat{\boldsymbol{\Sigma}}_{t_2}$ are finite by Lemma 6(a) and successive conditioning.

(a) The mean value theorem implies that $n(1) - N = n'(\varepsilon)$ for some ε between 0 and 1, so by Lemma 3(c),

$$|t_3(l) - N| \le 1 + \big(\tilde{\lambda}/\underline{\lambda}\big)N\big\|(1 + l/t_2)\hat{\boldsymbol{\Sigma}}_{t_2} - \mathbf{I}\big\|_\infty.$$

Now $(1 + l/t_2)\hat{\boldsymbol{\Sigma}}_{t_2}$ converges almost surely to the identity matrix since $\hat{\boldsymbol{\Sigma}}_m \to \mathbf{I}$ almost surely and since t_2 is defined to be larger than m. Also by Lemma 3, $(\tilde{\lambda}/\underline{\lambda}) \le (1 + \|(1 + l/t_2)\hat{\boldsymbol{\Sigma}}_{t_2} - \mathbf{I}\|_\infty)(1 + \|\hat{\boldsymbol{\Sigma}}_{t_2}^{-1} - \mathbf{I}\|_\infty)$, so $|t_3(l) - N|/N \to 0$ almost surely.

(b) Lemma 3 also implies that $(t_3(l)/N)^q$ is dominated by $(1 + \tilde{\lambda})^q$ for $q > 0$ and by $1 + (1 + l/m)^q(1 + \|\hat{\boldsymbol{\Sigma}}_{t_2}^{-1} - \mathbf{I}\|_\infty)^q$ for $q < 0$. The expectations of both dominating functions are finite by Lemma 5(a), so by part (a) and dominated convergence, $E[(t_3(l)/N)^q] \to 1$ as $N \to \infty$.

(c) The proofs of parts (c) and (d) use the following third-order Taylor series expansion of $\varkappa$ about 0,

$$(3.12)\quad t_3(l) - N = (t_3(l) - \varkappa(1)) + \varkappa'(0) + \left(\tfrac{1}{2}\right)\varkappa''(0) + \left(\tfrac{1}{6}\right)\varkappa'''(\varepsilon),$$

where ε is between 0 and 1.

Define $H(x) \equiv P\{\varkappa(1) \le x\} = P\{f(x,\mathbf{W}) \ge \gamma\}$. From Hall (1981), the expectation of $[\![\varkappa(1)]\!] - \varkappa(1)$ is

$$\tfrac{1}{2} - \int_0^1 \sum_{k=0}^{\infty} \int_{k+r}^{k+1} H''(x-r)\left(x-k-\tfrac{1}{2}\right) dx\, dr$$

and the integral does not exceed $\int_0^\infty |H''(x)|\,dx$ in absolute value. Applying a Helmert-type transformation, $\hat{\boldsymbol{\Sigma}}_n = (n-1)^{-1}\sum_{i=1}^{n-1}\mathbf{Y}_i\mathbf{Y}_i^T$, where $\mathbf{Y}_i = \{i(i+1)\}^{-1/2}(i\mathbf{X}_{i+1} - \sum_{j=1}^{i}\mathbf{X}_j)$, so that the $\mathbf{Y}_i$ are independent $\mathscr{N}(\mathbf{0},\mathbf{I})$ random vectors. Let $R = \{\mathbf{V}\colon\ \Phi(\mathbf{V}^{-1/2}A) \ge \gamma\}$. Then $H(x) = P\{\hat{\boldsymbol{\Sigma}}_{t_2}/x \in R\} = E[P\{(t_2-1)^{-1}\{\sum\mathbf{Y}_i\mathbf{Y}_i^T + (m-1)\hat{\boldsymbol{\Sigma}}_m\}/x \in R|\hat{\boldsymbol{\Sigma}}_m\}]$ and the conditional probability may be written using a Wishart $(t_2 - m - 1, \mathbf{I})$ distribution. By changing variables and differentiating, it is shown directly that $\int_0^\infty |H''(x)|\,dx = o(1)$, proving that $E[(t_3(l) - \varkappa(1))] = \frac{1}{2} + o(1)$.

Lemmas 3(b, g, h) and 5(a), together with (3.11), imply that

$$E\left[N_2^{1/2}|\varkappa'''(\varepsilon)|\right] \le N_2^{3/2}K_2E\left[\tilde{\lambda}^3\underline{\lambda}^{-5}\left\|(1+l/m)\hat{\boldsymbol{\Sigma}}_m - \mathbf{I}\right\|_\infty^3\right] = O(1),$$

so

$$E\left[|\varkappa'''(\varepsilon)|\right] = o(1).$$

Thus

$$E[t_3(l) - N] = \tfrac{1}{2} + E[\varkappa'(0)] + \left(\tfrac{1}{2}\right)E[\varkappa''(0)] + o(1).$$

$E[\varkappa''(0)]$ is evaluated explicitly in Lemma 5(c), so the proof of (c) is completed by evaluating, using Lemma 7,

$$(3.13)\quad E[\varkappa'(0)] = N\operatorname{tr}\left[\mathbf{M}E\left\{(1+l/t_2)\hat{\boldsymbol{\Sigma}}_{t_2} - \mathbf{I}\right\}\right] = l/c - 2[\operatorname{tr}\mathbf{M}^2]/c + o(1).$$

(d) We square terms in (3.12) to give

$$E\left[(t_3(l) - N)^2\right] = E\left[\{\varkappa'(0)\}^2 + \left(\tfrac{1}{4}\right)\{\varkappa''(\varepsilon)\}^2 + \varkappa'(0)\varkappa''(\varepsilon)\right].$$

$E[\{\varkappa'(0)\}^2]$ is shown to equal $2N\operatorname{tr}[\mathbf{M}^2]/c + o(N)$ in Lemma 5(b) and the other terms are shown to be $o(N)$ by applying Lemmas 3 and 5.

(e) By equation (3.12),

$$\sqrt{N_2}\,(t_3(l) - N)/N = \sqrt{N_2}\left[t_3(l) - \varkappa(1) + \varkappa'(0) + \left(\tfrac{1}{2}\right)\varkappa''(\varepsilon)\right]/N,$$

for some ε between 0 and 1. Now $\sqrt{N_2}\varkappa'(0)/N = \sqrt{N_2}\operatorname{tr}[(1+l/t_2)\hat{\boldsymbol{\Sigma}}_{t_2}\mathbf{M}]$ converges in distribution to a $\mathscr{N}(0, 2\operatorname{tr}[\mathbf{M}^2])$ random variable by Anscombe's (1952) theorem and $E[|\varkappa''(\varepsilon)|] = o(1)$ by Lemmas 3(f) and 5, so the limiting distribution of $\sqrt{N_2}\,(t_3(l) - N)/N$ is $\mathscr{N}(0, 2\operatorname{tr}[\mathbf{M}^2])$.

(f) Since $f(n,\mathbf{\Sigma}) = g(n,0)$, (3.1) may be rewritten as

$$P\{\overline{\mathbf{X}}_{t_3(l)} \in A\} = \gamma + g_1(N,0)E[t_3(l) - N] + (\tfrac{1}{2})g_{11}(N,0)E\left[(t_3(l) - N)^2\right] + (\tfrac{1}{2})E\left[\{g_{11}(n^*,0) - g_{11}(N,0)\}(t_3(l) - N)^2\right],$$

where n^* is between $t_3(l)$ and N. Using results (c), (d) and Lemma 4,

$$\begin{aligned} E[g(t_3(l),0)] &= \gamma + g_1(N,0)\{l/c - 2(\operatorname{tr}\mathbf{M}^2)/c + \tfrac{1}{2} + E[\varkappa''(0)]\} \\ &\quad + (\tfrac{1}{2})g_{11}(N,0)E\left[(\varkappa'(0))^2\right] \\ &\quad + (\tfrac{1}{2})E\left[\{g_{11}(n^*,0) - g_{11}(N,0)\}(t_3(l) - N)^2\right] + o(N^{-1}). \end{aligned}$$

The derivatives of g and the expected values of $[\varkappa'(0)]^2$ and $\varkappa''(0)$ are given in Lemmas 1 and 5. The proof is completed by showing that $E[|g_{11}(n^*,0) - g_{11}(N,0)|(t_3(l) - N)^2]$ is also $o(N^{-1})$. Because $\varkappa$ is continuous, $n^* = \varkappa(\varepsilon^*)$ for some ε^*. By the mean value theorem and Lemma 4(b, c),

$$\begin{aligned} &E\left[|g_{11}(n^*,0) - g_{11}(N,0)|(t_3(l) - N)^2\right] \\ &\quad \le E\left[\sup_{\varepsilon^*}|g_{111}(\varkappa(\varepsilon^*),0)|(t_3(l) - N)^2\right] \\ &\quad \le E\left[\sup_{\varepsilon^*}(2\varkappa(\varepsilon^*))^{-3}p\left[4p + (\varkappa(\varepsilon^*)/N)(\tilde{\lambda}/\underline{\lambda})K\right]^2(t_3(l) - N)^2\right], \end{aligned}$$

where the supremum is taken over all ε^* between 0 and 1. Thus $E[|g_{11}(n^*,0) - g_{11}(N,0)|(t_3(l) - N)^2] = o(N^{-1})$ by Lemmas 3, 4 and 7(a). This completes the proof of the theorem. □

PROOF OF COROLLARY 1. Using the mean value theorem and implicitly differentiating N,

$$\begin{aligned} &N\left[(1 + \hat{l}_3/t_2)\hat{\mathbf{\Sigma}}_{t_2}\right] - N\left[(1 + l_3/t_2)\hat{\mathbf{\Sigma}}_{t_2}\right] \\ &\quad = N\left[(1 + l_3^*/t_2)\hat{\mathbf{\Sigma}}_{t_2}\right][t_2 + l_3^*]^{-1}(\hat{l}_3 - l_3), \end{aligned}$$

for some l_3^* between $\hat{l}_3$ and l_3. Applying Lemma 3, then

$$\begin{aligned} &\left|N\left[(1 + \hat{l}_3/t_2)\hat{\mathbf{\Sigma}}_{t_2}\right] - N\left[(1 + l_3/t_2)\hat{\mathbf{\Sigma}}_{t_2}\right]\right| \\ &\quad \le (N/t_2)\left\|(1 + (\hat{l}_3 + l_3)/m)\hat{\mathbf{\Sigma}}_{t_2}\right\|_\infty |\hat{l}_3 - l_3|. \end{aligned} \tag{3.14}$$

We show that $|\hat{l}_3 - l_3|$ is small except on a set of small probability. Let $\delta = m^{-1/2}$ and $B = \{\|\hat{\mathbf{\Sigma}}_{t_2} - \mathbf{I}\|_\infty \le \delta\}$. Note that using Lemma 3(b) and the relationships between different matrix norms,

$$\left\|(t_3(0)/N)^{1/2}\hat{\mathbf{\Sigma}}_{t_2}^{1/2} - \mathbf{I}\right\|_\infty I_B \le 2\sqrt{p}\,\delta/(1 + \delta).$$

Thus the symmetric difference of the sets $\hat{A}'$ and A' is contained in $2\sqrt{p}\,\delta/(1 + \delta)A'$ on the set B. By Lemma 3(e), then,

$$\left\|\int_{\hat{A}'}\mathbf{x}\mathbf{x}^T\,d\Phi(\mathbf{x}) - \int_{A'}\mathbf{x}\mathbf{x}^T\,d\Phi(\mathbf{x})\right\|_\infty I_B \le 2\sqrt{p}\,\delta/(1 + \delta)K.$$

This result then implies that

$$|l_3 - \hat{l}_3| I_B \le r(\delta) \tag{3.15}$$

for $r(\delta)$ a nonrandom function of δ which tends to 0 as $\delta \to 0$.

We then show that $|l_3 + \hat{l}_3|$ is bounded by a function of Λ on B^c. By Stokes' theorem, Lemma 3(e) and Equation (2.3),

$$\begin{aligned} -\tfrac{1}{2}\big[E_{A'}[\mathbf{I}_x]\big]^{-1} & E_{A'}\big[2p - 4 + \mathbf{I}_x(\mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{M}\mathbf{x})\big] \\ &\le \big[E_{A'}[\mathbf{I}_x]\big]^{-1}\Big[\sum (-1)^{i-1}\int_{\partial A'} x_i \mathbf{x}^T\mathbf{M}\mathbf{x}\, d\Phi^{(i)}(\mathbf{x}) + 2E_{A'}[\mathbf{M}_x]\Big] \\ &\le \|\mathbf{M}\|_\infty K + 2\operatorname{tr}(\mathbf{M}^2) \\ &\le pK + 2. \end{aligned}$$

Similarly, using Lemma 3(e),

$$-\tfrac{1}{2}\big[E_{\hat{A}'}[\mathbf{I}_x]\big]^{-1}E_{\hat{A}'}\big[2p - 4 + \mathbf{I}_x(\mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\hat{\mathbf{M}}\mathbf{x})\big] \le pK\lambda_1\lambda_p^{-1} + 2.$$

Consequently,

$$(\hat{l}_3 + l_3) \le p + 6 + 2pK\lambda_1\lambda_p^{-1}. \tag{3.16}$$

Equations (3.14), (3.15) and (3.16) imply that

$$\begin{aligned} &\Big|N\big[(1 + \hat{l}_3/t_2)\hat{\boldsymbol{\Sigma}}_{t_2}\big] - N\big[(1 + l_3/t_2)\hat{\boldsymbol{\Sigma}}_{t_2}\big]\Big| \\ &\quad \le t_2^{-1}Nr_1(\delta) + t_2^{-1}N\tilde{\lambda}^2\underline{\lambda}^{-1}(p + 6 + 2pK)(1 + r_2(\delta))I_{B^c}, \end{aligned} \tag{3.17}$$

where $r_1(\delta)$ and $r_2(\delta)$ are nonrandom functions of δ which tend to 0 as $\delta \to 0$.

Inequality (3.17) and the fact that $P\{B^c\} = o(1)$ by Chebychev's inequality are then used to prove the corollary. The first-order asymptotic results of parts (a) and (b) follow immediately from (3.17) and Lemma 3.

To prove (c), note that

$$\begin{aligned} E\big[t_3(\hat{l}_3) - t_3(l_3)\big] &= E\big[t_3(\hat{l}_3) - N\big[(1 + \hat{l}_3/t_2)\hat{\boldsymbol{\Sigma}}_{t_2}\big]\big] \\ &\quad - E\big[t_3(l_3) - N\big[(1 + l_3/t_2)\hat{\boldsymbol{\Sigma}}_{t_2}\big]\big] \\ &\quad + E\big[N\big[(1 + \hat{l}_3/t_2)\hat{\boldsymbol{\Sigma}}_{t_2}\big] - N\big[(1 + l_3/t_2)\hat{\boldsymbol{\Sigma}}_{t_2}\big]\big]. \end{aligned}$$

The first two expectations on the right-hand side are both equal to $\frac{1}{2} + o(1)$ by the proof of Theorem 2(c). The Cauchy–Schwarz inequality, inequality (3.17) and Lemmas 5 and 7 imply that

$$\begin{aligned} &E\Big[\Big|N\big[(1 + \hat{l}_3/t_2)\hat{\boldsymbol{\Sigma}}_{t_2}\big] - N\big[(1 + l_3/t_2)\hat{\boldsymbol{\Sigma}}_{t_2}\big]\Big|\Big] \\ &\quad \le NE\big[t_2^{-1}\big[r_1(\delta) + \tilde{\lambda}^2\underline{\lambda}^{-1}(p + 6 + 2pK)(1 + r_2(\delta))I_{B^c}\big]\big] = o(1), \end{aligned}$$

proving (c).

It is similarly shown that

$$E\Big[\big(t_3(\hat{l}_3) - t_3(l_3)\big)^2\Big] = o(N) \tag{3.18}$$

and that $E[\sqrt{N_2}\,|t_3(\hat{l}_3) - t_3(l_3)|/N] \to 0$ as $N \to \infty$, proving parts (d) and (e).

Because the stopping rule t_3 is an increasing function of l, a first-order Taylor series expansion about the first argument gives

$$E\big[g\big(t_3(\hat{l}_3),0\big)\big] = E\big[g\big(t_3(l_3),0\big)\big] + E\big[g_1\big(t_3(l_3^*),0\big)\big(t_3(\hat{l}_3) - t_3(l_3)\big)\big],$$

where l_3^* is between l_3 and $\hat{l}_3$. Lemma 4(a), (3.17) and the Cauchy–Schwarz inequality imply that

$$E\big[g_1\big(t_3(l_3^*),0\big)\big|t_3(\hat{l}_3) - t_3(l_3)\big|\big] = o(N^{-1}).$$

Thus

$$E\big[g\big(t_3(\hat{l}_3),0\big)\big] = E\big[g\big(t_3(l_3),0\big)\big] + o(N^{-1}) = \gamma + o(N^{-1}),$$

by Theorem 2(f) and the definition of l_3. □

Acknowledgments. I would like to thank a referee and an Associate Editor for their careful reading and many helpful comments. I would also like to thank my thesis advisor, Mark Finster, for his guidance in this work.

REFERENCES

ANSCOMBE, F. J. (1952). Large-sample theory of sequential estimation. *Proc. Cambridge Philos. Soc.* **48** 600–607.

ANSCOMBE, F. J. (1953). Sequential estimation. *J. Roy. Statist. Soc. Ser. B* **15** 1–21.

CHATTERJEE, S. K. (1959). On an extension of Stein's two-sample procedure to the multinomial problem. *Calcutta Statist. Assoc. Bull.* **8** 121–148.

CHATTERJEE, S. K. (1960). Some further results on the multinomial extension of Stein's two-sample procedure. *Calcutta Statist. Assoc. Bull.* **9** 20–28.

CHOW, Y. S. and ROBBINS, H. (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist.* **36** 457–462.

COX, D. R. (1952). Estimation by double sampling. *Biometrika* **39** 217–227.

DANTZIG, G. B. (1940). On the non-existence of tests of "Student's" hypothesis having power functions independent of σ. *Ann. Math. Statist.* **11** 186–192.

FINSTER, M. P. (1985). Estimation in the general linear model when the accuracy is specified before data collection. *Ann. Statist.* **13** 663–675.

FINSTER, M. P. (1986). A theory of accurate estimation for multivariate data with applications to large-scale Monte Carlo studies. School of Business Working Paper, Univ. Wisconsin.

FISHMAN, G. S. (1977). Achieving specific accuracy in simulation output analysis. *Comm. ACM* **20** 310–315.

GRAYBILL, F. A. (1983). *Matrices with Applications in Statistics*. Wadsworth, Belmont, Calif.

HALL, P. (1981). Asymptotic theory of triple sampling for sequential estimation of a mean. *Ann. Statist.* **9** 1229–1238.

KLEIJNEN, J. P. C. (1984). *Statistical Tools for Simulation Practitioners*. Dekker, New York.

LAVENBERG, S. and SAUER, C. (1977). Sequential stopping rules for the regenerative method of simulation. *IBM J. Res. Develop.* **21** 545–558.

LOHR, S. (1988). Accurate multivariate estimation using double and triple sampling. Technical Report 505, School of Statist., Univ. Minnesota.

MUKHOPADHYAY, N. and AL-MOUSAWI, J. S. (1986). Fixed-size confidence regions for the mean vector of a multinormal distribution. *Sequential Anal.* **5** 139–168.

ROBBINS, H. (1959). Sequential estimation of the mean of a normal population. In *Probability and Statistics: The Harald Cramér Volume* (U. Grenander, ed.) 235–245. Almquist and Wiksell, Uppsala.

SIMONS, G. (1968). On the cost of not knowing the variance when making a fixed-width confidence interval for the mean. *Ann. Math. Statist.* **39** 1946–1952.

SPIVAK, M. (1965). *Calculus on Manifolds*. Benjamin, New York.

SRIVASTAVA, M. S. (1967). On fixed-width confidence bounds for regression parameters and mean vector. *J. Roy. Statist. Soc. Ser. B* **29** 132–140.

SRIVASTAVA, M. S. (1971). On fixed-width confidence bounds for regression parameters. *Ann. Math. Statist.* **42** 1403–1411.

SRIVASTAVA, M. S. and BHARGAVA, R. P. (1979). On fixed-width confidence region for the mean. *Metron* **37** 163–174.

STEIN (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16** 243–258.

U.S. ENVIRONMENTAL PROTECTION AGENCY (1982). Test methods for evaluating solid waste. U.S. EPA Publication SW–846, Office of Solid Waste and Emergency Response, U.S. EPA, Washington.

VARADHAN, S. R. S. (1984). *Large Deviations and Applications*. SIAM, Philadelphia.

WISHART, J. (1928). The generalized product moment distribution in samples from a normal multivariate population. *Biometrika* **20A** 32–52.

WOODROOFE, M. (1977). Second order approximations for sequential point and interval estimation. *Ann. Statist.* **5** 984–995.

WORKING, H. and HOTELLING, H. (1929). Application of the theory of error to the interpretation of trends. *J. Amer. Statist. Assoc. Supp.* (*Proc.*) **24** 73–85.

DEPARTMENT OF MATHEMATICS
ARIZONA STATE UNIVERSITY
TEMPE, ARIZONA 85287-1804