ACCURATE MULTIVARIATE ESTIMATION USING DOUBLE AND TRIPLE SAMPLING

Sharon L. Lohr[1]
University of Minnesota
Technical Report No. 505
February, 1988

(Abbreviated Title: ACCURATE MULTIVARIATE ESTIMATION)

AMS1980 subject classifications: Primary 62H12; secondary 62F25, 62L12

Key words and phrases: Two-stage estimation, three-stage estimation, confidence sets, multivariate normal distribution, sequential methods

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

ACCURATE MULTIVARIATE ESTIMATION USING DOUBLE AND TRIPLE SAMPLING

Summary

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. Triple-sampling procedures, 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 provide correction factors which make the procedures asymptotically consistent. Double sampling and triple sampling are both asymptotically efficient; in addition, the regret for triple sampling is a bounded function of the covariance structure and is independent of m.

1. Introduction

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ 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 t such that the resulting estimator $\hat{\boldsymbol{\theta}}_t$ accurately estimates $\boldsymbol{\theta}$. Accurate estimation is used here in the sense of Finster (1985, 1987). A fixed-accuracy set is a natural extension of a fixed-width confidence interval to $\mathbb{R}^p$: $\hat{\boldsymbol{\theta}}$ accurately estimates $\boldsymbol{\theta}$ with accuracy A and confidence γ if $P(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \in A) \geqq \gamma$. Formally, a fixed-accuracy set is a compact, orientable Borel-measurable set $A \in \mathbb{R}^p$ 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 (Office of Solid Waste and Emergency Response (1982), p. 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 i^{th} contaminant is estimated to 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 in the multivariate normal situation with unknown covariance, a data collection procedure which collects a fixed number of observations can not guarantee a desired predetermined accuracy and confidence level; a sequential or step-sequential procedure is therefore necessary. In many cases, however, a purely sequential procedure, in which the parameters are re-estimated after each observation, is impractical because of a delayed response or a difficulty in setting up the experiment, or even because the sample-size saving is not worth the inconvenience of repeated statistical analysis. Following Stein (1945), Cox (1952), and Hall (1981), who studied the one-dimensional case of fixed-width confidence interval estimation, we use double and triple sampling to limit data collection to two or three stages.

If $\boldsymbol{\Sigma}$ were known and the population were normal, any sample size n exceeding the solution N of $f(N,\boldsymbol{\Sigma}) = \gamma$, where

$$\begin{aligned} f(n,\mathbf{V}) &\equiv P\{\mathcal{N}(\mathbf{0},n^{-1}\mathbf{V}) \in A\} \\ &= \int_A (n/2\pi)\,|\mathbf{V}|^{-1/2}\exp[-(n/2)\,\mathbf{x}^T\mathbf{V}^{-1}\mathbf{x}]\,d\mathbf{x} \qquad (1.1) \end{aligned}$$

and $\mathcal{N}(\mathbf{0},n^{-1}\mathbf{V})$ represents a random vector with that distribution, would ensure that $\overline{\mathbf{X}}_n$ is an accurate estimator of $\boldsymbol{\theta}$. For $\boldsymbol{\Sigma}$ unknown, the double- and triple-sampling procedures of this paper both prescribe collecting a first sample of size m and estimating Σ by

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

A natural estimator of N after the pilot sample has been collected is $\hat{N}$, the solution $\hat{N}$ to $f(\hat{N}, \hat{\Sigma}_m) = \gamma$. Theorem 2 will show that $\hat{N}$ is asymptotically unbiased; however, the coverage probability using $\hat{N}$ is strictly less than γ up to $o(m^{-1})$ terms. Intuitively, the probability is less than γ because only a fraction of the data are used to estimate Σ: the conditional distribution of $\hat{\Sigma}_m^{-1/2}\overline{X}$ (the normal distribution) is used to find $\hat{N}$ while the actual distribution of $\hat{\Sigma}_m^{-1/2}\overline{X}$ 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.

The double-sampling stopping rule used here to give an asymptotically consistent procedure inflates the covariance estimate by a factor $(1+\ell/m)$ to compensate for not knowing Σ. The stopping rule for the double-sampling procedure, $\tau(\ell)$, is then the smallest integer n for which $f(n, (1+\ell/m)\hat{\Sigma}_m) \geq \gamma$. The parameter ℓ can be chosen so that the stopping rule $\tau(\ell)$ gives coverage probability γ with error $o(m^{-1})$.

If m is small relative to N, however, the double-sampling procedure will be inefficient when compared with the purely sequential procedures of

Chow and Robbins (1965) and Woodroofe (1977) for one-dimensional accurate estimation and Finster (1987) for multi-dimensional accurate estimation. The triple-sampling procedure 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 Σ were known, is set equal to $[[c(N-m)]] + m$, where $[[x]]$ denotes the smallest integer containing x. Then N_2 is estimated after the pilot sample by the stopping time

$$t_2 = [[c(\tau(0)-m)_+]] + m,$$

where $\tau(0)$ is the double-sampling stopping rule. After the second sample of t_2-m observations, the covariance matrix is re-estimated using $\hat{\Sigma}_{t_2}$, the least squares estimate of Σ using all t_2 observations. Then the size of the third and final sample is $[[t_3(\ell)]] - t_2$, where $t_3(\ell)$ satisfies the relation $f(t_3(\ell), (1 + \ell/t_2)\hat{\Sigma}_{t_2}) = \gamma$. Here ℓ is again a correction for the sequential nature of the procedure, giving the bounded regret of Simons (1968). With ℓ defined in (3.2), the triple-sampling procedure has finite regret and achieves coverage probability γ with error $o(N^{-1})$, the same order obtained by Finster (1987). With this small order of error, the asymptotic results for the triple-sampling procedure are valid even for 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

$$P\{\,|\mathbf{c}^T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})| \leq d\,\|\mathbf{c}\|_{q'},\ \forall\, \mathbf{c} \in \mathbb{R}^p\,\} \geq P\{\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|_q \leq d\} \geq \gamma. \tag{1.2}$$

The values q=2 and q=∞ give fixed-accuracy analogues of Scheffé's and Tukey's procedures for obtaining simultaneous confidence intervals. See Miller (1978).

The definition of accuracy used in this paper is that given by Finster (1985, 1987). The techniques used to develop the asymptotic properties for double and 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.

The double-sampling procedures are derived in section 2, and the triple-sampling procedures are derived and compared with Finster's (1987) purely sequential procedure in section 3. Section 4 contains the proofs of the main results.

2. Double-sampling procedures for accurate estimation

The goal of accurate estimation is to find an efficient stopping rule t for which $P(\overline{\mathbf{X}}_t - \boldsymbol{\theta} \in A) \geqq \gamma$ for a given accuracy set A and confidence coefficient γ. 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 for the double- and triple-sampling procedures are expressed in terms of N increasing to infinity. Note that N increases to infinity either as $\boldsymbol{\Sigma} \to \infty$ or as the accuracy set decreases to the empty set. We take "$\boldsymbol{\Sigma} \to \infty$ as $N \to \infty$" to mean that

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

is a constant set as $N \to \infty$. In other words, $\boldsymbol{\Sigma} \to \infty$ 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 (1.1) is replaced by dA, then $N \to \infty$ as $d \to 0$.

If one were to perform a double-sampling experiment and had a rough idea of the sample size needed to achieve the desired accuracy and coverage probability, one would typically take, say, half of the observations in the pilot sample. If the *a priori* estimate of N were much greater than the actual sample size needed, one would not have wasted too many observations; on the other hand, an underestimate of N could be corrected after the pilot sample. Much of the previous double-sampling work implicitly assumes that the pilot-sample size tends to infinity at the same rate as N, so that the resulting double-sampling procedure is asymptotically efficient. We make the weaker assumption that the pilot-sample size m tends to infinity as a fractional power of N, so that $N = O(m^h)$ for some $h \geq 1$, enabling us to determine the convergence rates for the "worst-case" situation in which the pilot-sample size is very small relative to N.

In the definition of the double-sampling procedure, $\tau(\ell) = [[N((1+\ell/m)\hat{\mathbf{\Sigma}}_m)]]$, where the sample size function $N(\mathbf{V})$ is defined as the solution to

$$f(N(\mathbf{V}),\mathbf{V}) = \gamma. \tag{2.2}$$

The following theorem demonstrates that conditionally on the stopping time $\tau(\ell)$ the parameter estimate $\overline{\mathbf{X}_{\tau(\ell)}}$ is normally distributed with mean $\mathbf{\Theta}$ and covariance $\mathbf{\Sigma}/\tau(\ell)$. In other words, conditionally on $\tau(\ell)$, $\overline{\mathbf{X}_{\tau(\ell)}}$ has the same distribution it would have if $\tau(\ell)$ were a fixed integer rather than a random variable. Thus $\overline{\mathbf{X}_{\tau(\ell)}}$ is unbiased. The proof of theorem follows the proofs of lemmas 1 through 4 in Robbins (1959).

Theorem 1. Let t be an integer-valued stopping time which is a function of $\{\hat{\mathbf{\Sigma}}_{p+1}, \hat{\mathbf{\Sigma}}_{p+2}, \hat{\mathbf{\Sigma}}_{p+3}, \ldots\}$. Then

(a) t is independent of $\overline{\mathbf{X}_k}$ for all k.

(b) The conditional distribution of $\overline{\mathbf{X}_t}$ given $t = n$ is $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}/n)$.

We now state the main result about second-order properties of the double-sampling procedure. Throughout, let Φ represent the standard multivariate normal probability measure and define

$$\mathfrak{M} \equiv \int_{A'} (\mathbf{I} - \mathbf{x}\mathbf{x}^T)\, d\Phi(\mathbf{x}) \left[\int_{A'} (p - \mathbf{x}^T\mathbf{x})\, d\Phi(\mathbf{x}) \right]^{-1}. \tag{2.3}$$

Theorem 2. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be independent and identically distributed

$\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ random vectors. Let $N = N(\boldsymbol{\Sigma})$ and $\tau(\ell) = [[N((1+\ell/m)\hat{\boldsymbol{\Sigma}}_m)]]$, where ℓ is a known constant and the function $N(\mathbf{V})$ is defined in (2.2). Assume $m \to \infty$ as a fractional power of N, so that $N = O(m^h)$ for some $h \geqq 1$. Then, as $N \to \infty$,

(a) $\tau(\ell)/N \to 1$ almost surely.

(b) For any $q \in \mathcal{R}$, $E\{[\tau(\ell)/N]^q\} \to 1$.

(c)
$$E[\tau(\ell)] = N + N\ell/m + 1/2$$
$$+ (N/2m)\left[\int_{A'} [p - \mathbf{x}^T\mathbf{x}]\, d\Phi(\mathbf{x})\right]^{-1} \left[\int_{A'} \{2\,p\,(\mathrm{tr}\,\mathfrak{M}^2) - 4\,2p\right.$$
$$\left. + (p - \mathbf{x}^T\mathbf{x})[\,\mathbf{x}^T\mathbf{x}\,(\mathrm{tr}\,\mathfrak{M}^2 + 1) - 2\,\mathbf{x}^T\mathfrak{M}\mathbf{x}]\}\, d\Phi(\mathbf{x})\right]$$
$$+ (N/2m)\,[(4 - p)\,\mathrm{tr}(\mathfrak{M}^2) - p - 2] + o(N/m).$$

(d) $E[(\tau(\ell) - N)^2] = 2\,N^2\,\mathrm{tr}\,(\mathfrak{M}^2)/m + o(N^2/m)$.

(e) $\sqrt{m}\,(\tau(\ell) - N)/N$ converges to a $\mathcal{N}(0, 2\,\mathrm{tr}(\mathfrak{M}^2))$ distribution.

(f)
$$P\{\overline{\mathbf{X}}_{\tau(\ell)} - \boldsymbol{\theta} \in A\} = \gamma + (4m)^{-1}\Big\{[2\,\ell + 4\,\mathrm{tr}(\mathfrak{M}^2) - p - 2]\left[\int_{A'} [p-\mathbf{x}^T\mathbf{x}]\, d\Phi(\mathbf{x})\right]$$
$$+ \int_{A'} \{-4 + 2p + (p - \mathbf{x}^T\mathbf{x})[\,\mathbf{x}^T\mathbf{x} - 2\,\mathbf{x}^T\mathfrak{M}\mathbf{x}]\}\, d\Phi(\mathbf{x})\Big\} + o(m^{-1}).$$

To attain asymptotically correct coverage probability up to $o(m^{-1})$ terms, we find the value ℓ_2 which solves $P\{\overline{\mathbf{X}}_{\tau(\ell)} - \boldsymbol{\theta} \in A\} = \gamma + o(m^{-1})$. Set

$$\ell_2 \equiv -\left[2\int_{A'} [p - \mathbf{x}^T\mathbf{x}]\, d\Phi(\mathbf{x})\right]^{-1}\left[\int_{A'} \{2p - 4 + (p - \mathbf{x}^T\mathbf{x})[\mathbf{x}^T\mathbf{x} - 2\,\mathbf{x}^T\mathfrak{M}\mathbf{x}]\}\, d\Phi(\mathbf{x})\right]$$
$$- 2\,\mathrm{tr}\,(\mathfrak{M}^2) + p/2 + 1. \qquad (2.4)$$

The first two terms in (2.4) depend upon N and $\boldsymbol{\Sigma}$ through the set A'. We substitute the estimates $\tau(0)$ and $\hat{\boldsymbol{\Sigma}}_m$ for N and $\boldsymbol{\Sigma}$ in (2.4) and define

$$\hat{A}' \equiv (\tau(0)\,\hat{\boldsymbol{\Sigma}}_m^{-1})^{1/2} A,$$

$$\hat{\mathfrak{M}} \equiv \int_{\hat{A}'} (\mathbf{I} - \mathbf{x}\mathbf{x}^T)\, d\Phi(\mathbf{x}) \left[\int_{\hat{A}'} (p - \mathbf{x}^T\mathbf{x})\, d\Phi(\mathbf{x})\right]^{-1},$$

and

$$\hat{\ell}_2 \equiv -[2\int_{\hat{A}'}[p - \mathbf{x}^T\mathbf{x}]\,d\Phi(\mathbf{x})]^{-1}[\int_{\hat{A}'}\{2p - 4 + (p - \mathbf{x}^T\mathbf{x})[\mathbf{x}^T\mathbf{x} - 2\,\mathbf{x}^T\hat{\mathfrak{M}}\mathbf{x}]\}\,d\Phi(\mathbf{x})]$$
$$- 2\,\mathrm{tr}\,(\hat{\mathfrak{M}}^2) + p/2 + 1. \qquad (2.5)$$

The following corollary to theorem 2 states that the results of the theorem hold when ℓ_2 is replaced by the random variable $\hat{\ell}_2$.

Corollary 1. Let $\tau(\hat{\ell}_2) = [[N((1 + \hat{\ell}_2/m)\hat{\Sigma}_m)]]$, where $\hat{\ell}_2$ is defined in (2.5) and N is defined in (2.2). Then the results of theorem 2 hold when $\hat{\ell}_2$ is substituted for ℓ. In particular,

(d) $$E[\tau(\hat{\ell}_2)] = N - p\,N\,\mathrm{tr}(\mathfrak{M}^2)/(2m)$$
$$+ (N/2m)[\int_{A'}[p - \mathbf{x}^T\mathbf{x}]\,d\Phi(\mathbf{x})]^{-1}[\int_{A'}\{2\,p\,(\mathrm{tr}\,\mathfrak{M}^2)$$
$$+ (p - \mathbf{x}^T\mathbf{x})\,\mathbf{x}^T\mathbf{x}\,\mathrm{tr}\,\mathfrak{M}^2\}\,d\Phi(\mathbf{x})] + o(N/m)$$

and

(f) $$P\{\overline{\mathbf{X}}_{\tau(\hat{\ell}_2)} - \boldsymbol{\theta} \in A\} = \gamma + o(m^{-1}).$$

We see from the theorem and corollary that the usual first-order asymptotic properties of pointwise and momentwise efficiency and asymptotic normality hold for the stopping rule $\tau(\ell)$. If m is very small relative to N, though, the stopping time has infinite regret and large variance; in addition, for small sample sizes the distribution of $\tau(\ell)$ is positively skewed because $\hat{\Sigma}_m$ has a Wishart distribution.

The factor $\mathrm{tr}\,(\mathfrak{M}^2)$ appears in the expressions for the variance of the stopping times. The matrix $\mathfrak{M}$ defined in (2.3) shows the effect of the shape and orientation of the standardized accuracy set $A' = (N\Sigma^{-1})^{1/2}A$ on the stopping times. From theorem 9.1.25 of Graybill (1983),

$$p^{-1} \leq \mathrm{tr}\,(\mathfrak{M}^2) \leq 1. \qquad (2.6)$$

We can get some feel for the meaning of $\mathrm{tr}\,(\mathfrak{M}^2)$ by examining the special case in which A is a spherical accuracy set. If A is spherical and if $\boldsymbol{\Sigma}$ is a diagonal matrix (i.e., the components of $\mathbf{X}$ are independent), then $\mathfrak{M}$ is also diagonal and hence $2\,\mathrm{tr}\,(\mathfrak{M}^2) = 2/p$. On the other hand, 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, $2\,\mathrm{tr}\,(\mathfrak{M}^2)$ will be close to two, the variance for the one-dimensional procedure of Cox (1952).

The theorem and corollary are proven in section 4, assuming throughout the proof without loss of generality that $\boldsymbol{\theta}$ is the zero vector and $\boldsymbol{\Sigma}$ is the p×p identity matrix. The method used to find the coverage probabilities and expected values, variances, and asymptotic distributions of the stopping rules relies on a Taylor series expansion of $N((1+\ell/m)\hat{\boldsymbol{\Sigma}}_m)$ about $\boldsymbol{\Sigma}$, using Fréchet derivatives. The Fréchet derivatives guarantee that all matrices will be positive definite.

The independence of $\tau(\ell)$ and $\overline{\mathbf{X}}_k$ allows the coverage probability to be calculated using the function f, defined in (1.1), as is shown in the following lemma.

Lemma 1. Let t be an integer-valued stopping time which is independent of $\overline{\mathbf{X}}_k$ for all k. Then

$$P[\,\overline{\mathbf{X}}_t \in A] = E[f(t, \boldsymbol{\Sigma})].$$

Proof.

$$P\{\mathbf{X}_t \in A\} = \sum_{k=1}^{\infty} P\{\mathbf{X}_k \in A, t = k\} = \sum_{k=1}^{\infty} f(k,\boldsymbol{\Sigma})\, P\{t = k\} = E[f(t,\boldsymbol{\Sigma})].$$

The second equality uses the independence of $\overline{\mathbf{X}_k}$ and t. □

By virtue of Lemma 1, then, the coverage probability is evaluated using the moments of $\tau(\ell)$.

$$P\{\overline{\mathbf{X}}_{\tau(\ell)} - \boldsymbol{\theta} \in A\} = E[f(\tau(\ell),\boldsymbol{\Sigma})]$$

$$= f(N,\boldsymbol{\Sigma}) + f_1(N,\boldsymbol{\Sigma})\, E[\tau(\ell)-N] + (1/2)\, E[f_{11}(n^*,\boldsymbol{\Sigma})\,(\tau(\ell)-N)^2]. \qquad (2.7)$$

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

The $O(m^{-1})$ terms in the expression for the coverage probability in theorem 2(f) result from substituting the sample covariance for the "true" covariance when determining the stopping time, without accounting for this substitution. For the one-dimensional case of estimating a mean, $f_{11}(N,\boldsymbol{\Sigma})$ is simply the first-order term in the Taylor series expansion of the $(1-\gamma)^{th}$ percentile of a t-distribution with m degrees of freedom about the $(1-\gamma)^{th}$ percentile of the normal distribution.

3. Triple sampling for accurate multivariate estimation

The double-sampling stopping rules of section two work very well if the pilot-sample size m has the same order of magnitude as the optimal sample size N. If $m/N \to 0$, however, the stopping time $\tau(\ell)$ leads to an inefficient procedure. The triple-sampling procedure achieves finite regret and second-order asymptotic efficiency by taking two additional samples after the pilot sample rather than just one.

In terms of the function $N(\mathbf{V})$ defined in (2.2), $t_3(\ell) = N[\,(1 + \ell/t_3)\,\hat{\boldsymbol{\Sigma}}_{t_2}\,]$. Since $t_3(\ell)$ is an implicit function of $\{\hat{\boldsymbol{\Sigma}}_{p+1}, \hat{\boldsymbol{\Sigma}}_{p+2}, \ldots\}$, Theorem 1 implies that conditionally on the stopping time $t_3(\ell)$, the parameter estimates $\overline{\mathbf{X}}_{t_3(\ell)}$ are normally distributed with mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma}/t_3(\ell)$ and hence are unbiased. We now state the second-order asymptotic properties of the triple-sampling procedure.

Theorem 3. Let $N = N(\boldsymbol{\Sigma})$, $t_2 = [[c(\,\tau(\ell) - m)]]_+ + m$, and $t_3(\ell) = N[(1+\ell/t_3)\,\hat{\boldsymbol{\Sigma}}_{t_2}\,]$, where ℓ is a known constant and the function N is defined in (2.2). 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) $t_3(\ell)/N \to 1$ almost surely.

(b) For any $q \in \mathbb{R}$, $E\{\,[\,t_3(\ell)/N]^q\,\} \to 1$.

(c) $$E[\,t_3(\ell)] = N + \ell/c - 2\,\mathrm{tr}\,\mathfrak{M}^2/\,c - 1/(2\,c) + 1/2 + (2c)^{-1}\,[\,\int_{A'} [p - \mathbf{x}^T\mathbf{x}]\, d\Phi(\mathbf{x})]^{-1}\,[\int_{A'} \{2\,p\,(\mathrm{tr}\,\mathfrak{M}^2) - 4 + 2p + (p - \mathbf{x}^T\mathbf{x})[\,\mathbf{x}^T\mathbf{x}\,(\mathrm{tr}\,\mathfrak{M}^2 + 1) - 2\,\mathbf{x}^T\mathfrak{M}\mathbf{x}]\,\}\, d\Phi(\mathbf{x})]$$

$$+ (2c)^{-1} [(4-p)\,\mathrm{tr}(\mathfrak{M}^2) - p - 2] + o(1).$$

(d) $E[(t_3(\ell) - N)^2] = 2N\,\mathrm{tr}(\mathfrak{M}^2)/c + o(N)$.

(e) $\sqrt{c}(t_3(\ell) - N)/\sqrt{N}$ converges to a $\mathcal{N}(0, 2\,\mathrm{tr}(\mathfrak{M}^2))$ distribution.

(f) $P\{\overline{\mathbf{X}}_{[[t_3(\ell)]]} - \boldsymbol{\theta} \in A\} = \gamma$

$$+ (4cN)^{-1}\{[2\ell - p - c - 2][\textstyle\int_{A'} [p - \mathbf{x}^T\mathbf{x}]\, d\Phi(\mathbf{x})]$$

$$+ \textstyle\int_{A'} [-4 + 2p + (p - \mathbf{x}^T\mathbf{x})[\mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathfrak{M}\mathbf{x}]]\, d\Phi(\mathbf{x})\} + o(N^{-1}).$$

To attain asymptotically correct coverage probability up to $o(N^{-1})$ terms, set

$$\ell_3 \equiv \ell_2 + 2\,\mathrm{tr}(\mathfrak{M}^2) - c/2. \tag{3.1}$$

We substitute the estimates $t_3(0)$ and $\hat{\boldsymbol{\Sigma}}_{t_2}$ for N and $\boldsymbol{\Sigma}$ in (3.1) and define

$$\tilde{A}' \equiv (t_3(0)\, \hat{\boldsymbol{\Sigma}}_{t_2}^{-1})^{1/2} A,$$

$$\tilde{\mathfrak{M}} \equiv \int_{\tilde{A}'} (\mathbf{I} - \mathbf{x}\mathbf{x}^T)\, d\Phi(\mathbf{x}) \left[\int_{\tilde{A}'} (p - \mathbf{x}^T\mathbf{x})\, d\Phi(\mathbf{x})\right]^{-1},$$

and

$$\hat{\ell}_3 \equiv -[2\textstyle\int_{\tilde{A}'} [p - \mathbf{x}^T\mathbf{x}]\, d\Phi(\mathbf{x})]^{-1}[\textstyle\int_{\tilde{A}'} (2p - 4 + (p - \mathbf{x}^T\mathbf{x})[\mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\tilde{\mathfrak{M}}\mathbf{x}])\, d\Phi(\mathbf{x})]$$

$$+ p/2 + 1 - c/2. \tag{3.2}$$

The following corollary to Theorem 3 states that the results of Theorem 3 hold when ℓ_3 is replaced by the random variable $\hat{\ell}_3$. The proof of the corollary is similar to that of corollary 1 and is omitted here.

Corollary 2. Let $t_3(\hat{\ell}_3) = N[(1 + \hat{\ell}_3/t_3)\, \hat{\boldsymbol{\Sigma}}_{t_2}]$, where $\hat{\ell}_3$ is defined in (3.2). Suppose the conditions of theorem 3 hold. Then, as $N \to \infty$,

(a) $t_3(\hat{\ell}_3)/N \to 1$ almost surely.

(b) For any $q \in \mathcal{R}$, $E\{[t_3(\hat{\ell}_3)/N]^q\} \to 1$.

(c) $E[t_3(\hat{\ell}_3)] = N - p\,\mathrm{tr}(\mathfrak{M}^2)/2c + 1/2$

$$+ (2c)^{-1}\left[\int_{A'} [p - \mathbf{x}^T\mathbf{x}]\, d\Phi(\mathbf{x})\right]^{-1}\left[\int_{A'} \{2\,p\,(\mathrm{tr}\,\mathfrak{M}^2) + (p - \mathbf{x}^T\mathbf{x})\,\mathbf{x}^T\mathbf{x}\,\mathrm{tr}\,\mathfrak{M}^2\}\, d\Phi(\mathbf{x})\right] + o(1).$$

(d) $E[(t_3(\hat{\ell}_3) - N)^2] = 2N\,\mathrm{tr}(\mathfrak{M}^2)/c + o(N)$.

(e) $\sqrt{c}\,(t_3(\hat{\ell}_3) - N)/N$ converges to a $\mathcal{N}(0, 2\,\mathrm{tr}(\mathfrak{M}^2))$ distribution.

(f) $P\{\overline{\mathbf{X}}_{[[t_3(\hat{\ell}_3)]]}\} = \gamma + o(N^{-1})$.

Theorems 2 and 3 demonstrate that the double-sampling and triple-sampling procedures both attain 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 .95 or .99, 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 $O(m^{-1})$ in the expression for the coverage probability in Theorem 2(f).

The extra term $-2\,\mathrm{tr}\,(\mathfrak{M}^2)$ in the expression for the average sample number of the triple-sampling procedure (theorem 3(c)) shows the effect of optional stopping and appears because $\hat{\Sigma}_{t_2}$ has bias $-2\,\Sigma^{1/2}\,\mathfrak{M}\Sigma^{1/2}/N_2$. This bias is proven in lemma 11 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 overestimate 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 underestimate Σ. The argument that $\hat{\Sigma}_{t_2}$ is biased also applies to $\hat{\Sigma}_{T(\ell)}$ and $\hat{\Sigma}_{t_3(\ell)}$. If one ignored the fact that these quantities are obtained sequentially, substituting $\hat{\Sigma}_{t_2}$ 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 one-dimensional results of Cox (1952) and Hall (1981) follow as special cases of the results in theorems 2 and 3. Let d be the desired half-width of the confidence interval and let z be the $(1-\gamma)/2$ critical point of the standard normal distribution. Then $A' = [-z, z]$ and $\mathfrak{M} = 1$. Evaluating the integrals in theorems 2 and 3, $E[T(\ell)] = N + (N/m)(\ell + 1/2)$, $E[t_3(\ell)] = (\ell - 3/2)/c + 1/2$, $f_1(N,0) = (\sqrt{2\pi}\,N)^{-1}\,z\exp[-z^2/2]$, and $f_{11}(N,0) = -(8\pi)^{-1/2}\,N^{-2}\,(z + z^3)\exp[-z^2/2]$. Thus the value of ℓ

making the coefficient of m^{-1} in theorem 2(f) vanish is $\ell_2 = (1+z^2)/2$ and the triple-sampling procedure which uses $\ell_3 = (1+z^2)/2 + 2 - c/2$ for ℓ will have coverage probability $\gamma + o(N^{-1})$. These are the results obtained by Cox and Hall.

Hall (1981) recommends using 1/2 for c. An alternative choice uses the distribution of $\tau(0)$. Since the distribution of $\tau(0)$ is approximately $\mathcal{N}(N, 2N^2 \operatorname{tr}(\mathfrak{M}^2)/m))$ and since $\operatorname{tr}(\mathfrak{M}^2) \leq 1$, $\tau(0)[1 - z_\alpha (2/m)^{1/2}]$ is an approximate $(1 - \alpha)$ lower confidence bound for N. This suggests taking c to be $1 - z_\alpha (2/m)^{1/2}$.

Table 3.1 compares the properties of these multivariate double- and triple-sampling procedures and Finster's (1987) purely sequential procedure. The quantity ℓ_2, which appears in all of the correction factors, is messy to calculate exactly but may be bounded by $p/2 + pK^2$, where K is the radius of the smallest sphere which will circumscribe the standardized accuracy set A', defined in (2.1). In practice, we may use the radius of the smallest sphere circumscribing the sample standardized accuracy set ($[\tau(0)\hat{\Sigma}_m^{-1}]^{1/2}A$ for double-sampling, $[t_3(0)\hat{\Sigma}_{t_2}^{-1}]^{1/2}A$ for triple-sampling) instead of K.

Table 3.1. Properties of the double-sampling, triple-sampling, and purely sequential procedures for multivariate estimation. Below, m is the size of the first sample, c is the fraction of observations taken in the second sample, N is the "best" fixed-sample stopping rule, and ρ corrects for the discreteness of the purely sequential stopping rule. Also, $\mathfrak{M}$ and ℓ_2 are defined in (2.3) and (2.4), and

$$r = (\mathrm{tr}\,\mathfrak{M}^2/2)\,\{-\rho + [\int_{A'} [p - \mathbf{x}^T\mathbf{x}]\,d\Phi(\mathbf{x})]^{-1}[\int_{A'} \{2p + (p - \mathbf{x}^T\mathbf{x})\mathbf{x}^T\mathbf{x}\}\,d\Phi(\mathbf{x})]\}.$$

	Double Sampling	Triple Sampling	Purely Sequential
Correction factor ℓ	ℓ_2	$\ell_2 + 2\,(\mathrm{tr}\,\mathfrak{M}^2)$	$\ell_2 + 2(\mathrm{tr}\,\mathfrak{M}^2) - \rho$
Regret	$r\,\mathrm{N/m}$	r/c	r
Asymptotic variance of stopping time	$2\,(\mathrm{tr}\,\mathfrak{M}^2)\,\mathrm{N}^2/\mathrm{m}$	$2\,(\mathrm{tr}\,\mathfrak{M}^2)\,\mathrm{N}/c$	$2\,(\mathrm{tr}\,\mathfrak{M}^2)\,\mathrm{N}$
Approximate distribution of stopping time	linear combination of χ^2_1	Normal	Normal
Coverage probability	$\gamma + o(\mathrm{m}^{-1})$	$\gamma + o(\mathrm{N}^{-1})$	$\gamma + o(\mathrm{N}^{-1})$

4. Proofs

The stopping rules for the double and triple-sampling procedures are of the form N(**W**), with $\mathbf{W} = (1 + \ell/m)\,\hat{\Sigma}_m$ for double sampling and $\mathbf{W} = (1 + \ell/t_2)\,\hat{\Sigma}_{t_2}$ for triple sampling. To calculate the Fréchet derivatives of N(**W**), used in the Taylor series expansion of N(**W**) about **I**, define the function

$$n(\epsilon) = N(\epsilon \mathbf{W} + (1-\epsilon)\,\mathbf{I}), \tag{4.1}$$

for positive definite **W** and $0 \leqq \epsilon \leqq 1$. The proofs of the theorems involve evaluating the first two derivatives of $n(\epsilon)$ and bounding the third. These derivatives are more easily evaluated and bounded in a different coordinate system. Let $\mathbf{\Lambda}$ be the matrix of eigenvalues of **W** and **P** the matrix of eigenvectors of **W**. Then

$$\epsilon \mathbf{W} + (1-\epsilon)\mathbf{I} = \mathbf{P}^T \mathbf{L}(\epsilon)\mathbf{P},$$

where

$$\mathbf{L}(\epsilon) = \epsilon\,[\mathbf{\Lambda} - \mathbf{I}] + \mathbf{I}. \tag{4.2}$$

Define the set

$$A^*(n,\epsilon) \equiv [n\,\mathbf{L}^{-1}(\epsilon)]^{1/2}\,\mathbf{P}^T A \tag{4.3}$$

and the functions

$$g(n,\epsilon) \equiv f(n, \epsilon\mathbf{W} + (1-\epsilon)\mathbf{I}) \tag{4.4}$$

for f defined in (1.1). Then $g(n,\epsilon)$ may be rewritten as

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

Lemma 2. Let $n(\epsilon)$ and $g(n,\epsilon)$ be defined in (4.1) through (4.4), and let **K** and **H** denote

$$\mathbf{K}(\epsilon) \equiv \mathbf{L}^{-1/2}(\epsilon)\,[\mathbf{\Lambda} - \mathbf{I}]\,\mathbf{L}^{-1/2}(\epsilon) \tag{4.5}$$

and

$$H(\epsilon) \equiv (n'(\epsilon)/n(\epsilon))\mathbf{I} - K(\epsilon), \tag{4.6}$$

respectively. Then

$$g_1(n,\epsilon) = (1/2n)\int_{A*(n,\epsilon)} [p - \mathbf{x}^T\mathbf{x}]\, d\Phi(\mathbf{x})$$

$$g_2(n,\epsilon) = -(1/2)\int_{A*(n,\epsilon)} [tr(K) - \mathbf{x}^T K\mathbf{x}]\, d\Phi(\mathbf{x})$$

$$g_{11}(n,\epsilon) = (1/4n^2)\int_{A*(n,\epsilon)} \{[p - \mathbf{x}^T\mathbf{x}]^2 - 2p\}\, d\Phi(\mathbf{x}).$$

$$g_{12}(n,\epsilon) = -(1/4n)\int_{A*(n,\epsilon)} \{[tr(K) - \mathbf{x}^T K\mathbf{x}][p - \mathbf{x}^T\mathbf{x}] - 2\mathbf{x}^T K\mathbf{x}\}\, d\Phi(\mathbf{x})$$

$$g_{22}(n,\epsilon) = (1/4)\int_{A*(n,\epsilon)} \{[tr(K) - \mathbf{x}^T K\mathbf{x}]^2 + 2tr[K^2] - 4\mathbf{x}^T K^2\mathbf{x}\}\, d\Phi(\mathbf{x})$$

$$g_{111}(n,\epsilon) = (8n^3)^{-1}\int_{A*(n,\epsilon)} [[p - \mathbf{x}^T\mathbf{x}]^3 - 6p\,[p - \mathbf{x}^T\mathbf{x}] + 8p]\, d\Phi(\mathbf{x})$$

$$n'(\epsilon) = n(\epsilon)\int_{A*(n(\epsilon),\epsilon)} [tr(K) - \mathbf{x}^T K\mathbf{x}]\, d\Phi(\mathbf{x}) \left\{\int_{A*(n(\epsilon),\epsilon)} [p - \mathbf{x}^T\mathbf{x}]\, d\Phi(\mathbf{x})\right\}^{-1}$$

$$n''(\epsilon) = [4g_1(n(\epsilon),\epsilon)]^{-1} \Big\{2\gamma tr(H^2) - \int_{A*(n(\epsilon),\epsilon)} [tr(H) - \mathbf{x}^T H\mathbf{x}]^2\, d\Phi(\mathbf{x})$$

$$+ 4\int_{A*(n(\epsilon),\epsilon)} [tr(HK) - \mathbf{x}^T HK\mathbf{x}]\, d\Phi(\mathbf{x})\Big\}$$

$$n'''(\epsilon) = [8g_1(n(\epsilon),\epsilon)]^{-1} \int_{A*(n(\epsilon),\epsilon)} \Big\{ tr\,[\,24\,\{n''(\epsilon)/n(\epsilon)\}\,H - 8H^3 - 24H^2K]$$

$$+12\,[tr(H) - \mathbf{x}^T H\mathbf{x}][tr(HK - \{n''(\epsilon)/n(\epsilon)\}\mathbf{I}) - \mathbf{x}^T(HK - \{n''(\epsilon)/n(\epsilon)\}\mathbf{I})\mathbf{x}]$$

$$+ 24\,[tr(\{n''(\epsilon)/n(\epsilon)\}K - HK^2) - \mathbf{x}^T(\{n''(\epsilon)/n(\epsilon)\}K - HK^2)\mathbf{x}]$$

$$- [tr(H) - \mathbf{x}^T H\mathbf{x}]^3 \Big\}\, d\Phi(\mathbf{x}).$$

Proof. The partial derivatives of g are found directly. Since $g(n(\epsilon),\epsilon) = \gamma$, applying the implicit function theorem gives the derivatives of $n(\epsilon)$. □

The partial derivative $g_1(n(\epsilon),\epsilon)$, appearing in the denominators of the expressions for $n'(\epsilon)$, $n''(\epsilon)$, and $n'''(\epsilon)$, is always greater than zero but can be small. To facilitate finding bounds for the derivatives of n, we apply Stokes's theorem to express them as integrals over the boundary of $A^* = A^*(n(\epsilon),\epsilon)$. Here, $dx^{(i)}$ is written for $dx_1 dx_2 \ldots dx_{i-1} dx_{i+1} \ldots dx_p$, ∂A^* represents the boundary of A^*, and $d\Phi^{(i)}(\mathbf{x}) = (2\pi)^{-p/2} \exp\{-\mathbf{x}^T\mathbf{x}/2\}\, dx^{(i)}$.

Lemma 3. Let $n(\epsilon)$, $g(n,\epsilon)$, $\mathbf{K}(\epsilon)$ and $\mathbf{H}(\epsilon)$ be defined in (4.1) through (4.6) and let $\mathbf{H}_i$ and $\mathbf{K}_i$ denote the i^{th} diagonal entries of $\mathbf{H}$ and $\mathbf{K}$.

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

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

(c) $$n'''(\epsilon) = [8g_1(n(\epsilon),\epsilon)]^{-1}$$

$$\Big\{-12 \sum (-1)^{i-1}\mathbf{H}_i \int_{\partial A^*} x_i\,[\mathbf{x}^T\mathbf{H}\mathbf{K}\mathbf{x} + 2\mathbf{H}_i\mathbf{K}_i]\, d\Phi^{(i)}(\mathbf{x})$$

$$+ 12 \sum (-1)^{i-1}\mathbf{H}_i\, \{n''(\epsilon)/n(\epsilon)\} \int_{\partial A^*} x_i\,[\mathbf{x}^T\mathbf{x} + 2]\, d\Phi^{(i)}(\mathbf{x})$$

$$+ 24 \sum (-1)^{i-1}\mathbf{K}_i\,(\{n''(\epsilon)/n(\epsilon)\} - \mathbf{H}_i\mathbf{K}_i) \int_{\partial A^*} x_i\, d\Phi^{(i)}(\mathbf{x})$$

$$- \sum (-1)^{i-1}\mathbf{H}_i \int_{\partial A^*} x_i[(\mathbf{x}^T\mathbf{H}\mathbf{x})^2$$

$$+ (4\mathbf{H}_i - 2\mathrm{tr}[\mathbf{H}])\, \mathbf{x}^T\mathbf{H}\mathbf{x} + 8\mathbf{H}_i^2 + 4\mathrm{tr}[\mathbf{H}^2] - 4\,\mathbf{H}_i\mathrm{tr}[\mathbf{H}]]\, d\Phi^{(i)}(\mathbf{x})\Big\}.$$

Proof. Let $\mathbf{C} = \mathrm{diag}(c_1, c_2, \ldots, c_p)$ and $\mathbf{D} = \mathrm{diag}(d_1, d_2, \ldots, d_p)$. Then by Stokes' theorem, quoted in Spivak (1965),

(i) $$\int_{A^*} [\mathrm{tr}(\mathbf{C}) - \mathbf{x}^T\mathbf{C}\mathbf{x}]\, d\Phi(\mathbf{x}) = \sum_{i=1}^{p} (-1)^{i-1} c_i \int_{\partial A^*} x_i\, d\Phi^{(i)}(\mathbf{x})$$

$$\text{(ii)} \int_{A*} [tr(\mathbf{C}) - \mathbf{x}^T\mathbf{C}\mathbf{x}]\, \mathbf{x}^T\mathbf{D}\mathbf{x}\, d\Phi(\mathbf{x}) = \sum_{i=1}^{p} (-1)^{i-1} c_i \int_{\partial A*} x_i\, \mathbf{x}^T\mathbf{D}\mathbf{x}\, d\Phi^{(i)}(\mathbf{x})$$

$$+ 2\sum_{i=1}^{p} (-1)^{i-1} c_i d_i \int_{\partial A*} x_i\, d\Phi^{(i)}(\mathbf{x}) - 2\gamma tr[\mathbf{DC}].$$

$$\text{(ii)} \int_{A*} [tr(\mathbf{C}) - \mathbf{x}^T\mathbf{C}\mathbf{x}]\, [\mathbf{x}^T\mathbf{C}\mathbf{x}]^2\, d\Phi(\mathbf{x}) = -8\gamma tr(\mathbf{C}^3) - 4\gamma tr(\mathbf{C}^2) tr(\mathbf{C})$$

$$+ \sum_{i=1}^{p} (-1)^{i-1} c_i \int_{\partial A*} x_i \left[[\mathbf{x}^T\mathbf{C}\mathbf{x}]^2 + 4c_i\, \mathbf{x}^T\mathbf{C}\mathbf{x} + 8c_i^2 + 4tr[\mathbf{C}^2] \right] d\Phi^{(i)}(\mathbf{x}).$$

The lemma follows by applying the identities above to the expressions for the derivatives in Lemma 2. □

All of the derivatives of n with respect to ϵ involve a linear combination of the integrals $\int_{A*} (1-x_i^2)\, d\Phi(\mathbf{x}) = (-1)^{i-1} \int_{\partial A*} x_i\, d\Phi^{(i)}(\mathbf{x})$ in the denominators. To aid in bounding the derivatives, we show in the following lemma that these integrals are positive.

Lemma 4. If $h(\mathbf{x}) > 0$ for all $\mathbf{x}$ and if R is an accuracy set, then

$$\int_R (1 - x_i^2)\, h(\mathbf{x})\, d\Phi(\mathbf{x}) > 0.$$

Proof. For $y > -1$ define

$$k(y) \equiv \int_R (2\pi)^{-p/2} h(\mathbf{x}) (1+y)^{1/2} \exp[-(\mathbf{x}^T\mathbf{x} + y x_i^2)/2]\, d\mathbf{x}.$$

$$= E[\, h(\mathbf{X})\, I_{(\mathbf{I}+\mathbf{Y})R}(\mathbf{X})\,],$$

where $\mathbf{X} \sim \mathcal{N}(\mathbf{0},\mathbf{I})$, $\mathbf{I}$ is the pxp identity matrix and $\mathbf{Y}$ is the pxp matrix with $(1+y)^{1/2} - 1$ in the $(i,i)^{th}$ entry and zeroes elsewhere, and I_B denotes the indicator variable of the event B. The set $(\mathbf{I}+\mathbf{Y})$ R increases with y because R is star-shaped with respect to zero. Since $h(\mathbf{X}) > 0$, k(y) strictly increases in y. Hence

$$2\,k'(0) = \int_R [(1+y)^{-1/2} - (1+y)^{1/2} x_i^2]\, h(\mathbf{x}) \exp[-yx_i^2/2]\, d\Phi(\mathbf{x}) \Big|_{y=0}$$

$$= \int_R (1 - x_i^2)\, h(\mathbf{x})\, d\Phi(\mathbf{x}) > 0. \quad \square$$

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

$$\tilde{\lambda} \equiv \max(\lambda_1, 1\}, \tag{4.7}$$

and

$$\underset{\sim}{\lambda} \equiv \min\{\lambda_p, 1\}, \tag{4.8}$$

where $\lambda_1 \geqq \lambda_2 \geqq \ldots \geqq \lambda_p$ are the eigenvalues of $\mathbf{W}$. For any matrix $\mathbf{C}$, let $\|\mathbf{C}\|_\infty$ represent the supremum norm of a matrix;

$$\|\mathbf{C}\|_\infty \equiv \sup \{ \|\mathbf{Cx}\|_\infty / \|\mathbf{x}\|_\infty\} = \max |C_{ij}|.$$

Lemma 5. Let $0 < \epsilon < 1$, and let $\mathbf{L}(\epsilon)$, $\mathbf{K}(\epsilon)$, $\mathbf{H}(\epsilon)$, $\tilde{\lambda}$, and $\underset{\sim}{\lambda}$ be as defined in (4.2) and (4.5) through (4.8). Then

(a) $\underset{\sim}{\lambda}\,\mathbf{I} \leqq \mathbf{L}(\epsilon) \leqq \tilde{\lambda}\,\mathbf{I}$.

(b) $\underset{\sim}{\lambda}\,N \leqq n(\epsilon) \leqq \tilde{\lambda}\,N$.

(c) $| n'(\epsilon)/ n(\epsilon) | \leqq \|\mathbf{K}(\epsilon)\|_\infty$.

(d) $\|\mathbf{H}(\epsilon)\|_\infty \leqq 2\,\|\mathbf{K}(\epsilon)\|_\infty$.

(e) There exists a constant K, independent of ϵ, such that $a \in A^*(N,0)$ implies $a^T a \leqq K$ and $a \in A^*(n(\epsilon),\epsilon)$ implies $a^T a \leqq (\tilde{\lambda}/\underset{\sim}{\lambda})\,K$.

(f) $|n''(\epsilon)| \leqq n(\epsilon)\,(\tilde{\lambda}/\underset{\sim}{\lambda})\,K_1\,\|\mathbf{K}(\epsilon)\|_\infty^2$, where $K_1 = 2p + 8 + 2K$.

(g) $|n'''(\epsilon)| \leqq n(\epsilon)\,\|\mathbf{K}(\epsilon)\|_\infty^3\,(\tilde{\lambda}/\underset{\sim}{\lambda})^2\,K_2$, where K_2 is a constant.

(h) $\|\mathbf{K}(\epsilon)\|_\infty \leqq \underset{\sim}{\lambda}^{-1}\,\|\mathbf{W} - \mathbf{I}\|$ for any matrix norm $\|\cdot\|$

Proof.

Part (a) follows immediately from the definition of $\mathbf{L}(\epsilon)$.

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

$$\gamma = \Phi[[n(\epsilon)\, \mathbf{L}^{-1}(\epsilon)]^{1/2}\mathbf{P}^{-1}A] \geq \Phi[[\mathbf{I}\, n(\epsilon)/\tilde{\lambda}]^{1/2}\mathbf{P}^{-1}A] > \Phi[N^{1/2}\mathbf{P}^{-1}A] = \gamma,$$

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

(c) From lemma 2, and using lemma 4 to show that each integral is positive,

$$n'(\epsilon) = n(\epsilon) \int_{A*} [\mathrm{tr}(\mathbf{K}(\epsilon)) - \mathbf{x}^T\mathbf{K}(\epsilon)\mathbf{x}]\, d\Phi(\mathbf{x}) \,[\, \int_{A*} [p - \mathbf{x}^T\mathbf{x}]\, d\Phi(\mathbf{x})\,]^{-1}$$

$$= n(\epsilon)\, [\, \sum \mathbf{K}_i(\epsilon) \int_{A*} [1 - x_i^2]\, d\Phi(\mathbf{x})]\, [\, \sum \int_{A*} [1 - x_i^2]\, d\Phi(\mathbf{x})\,]^{-1}$$

$$\leq n(\epsilon)\, \|\mathbf{K}(\epsilon)\|_\infty.$$

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

(e) Using (a), (c), and the definition of A',

$$A^*(n(\epsilon),\epsilon) = [\, n(\epsilon)/N\,]^{1/2}\, \mathbf{L}^{-1/2}(\epsilon)\, \mathbf{P}^T A' \subset (\tilde{\lambda}/\underset{\sim}{\lambda})^{1/2}\, \mathbf{P}^T A'.$$

Recall that the standardized accuracy set $A' = (N\, \Sigma^{-1})^{1/2}A$ is a constant, bounded set is thus contained in a ball of radius $\sqrt{K}$ for some $K < \infty$. Thus $a^T a \leq K$ if $a \in A^*(N,0)$, and $a^T a \leq (\tilde{\lambda}/\underset{\sim}{\lambda})\, K$ if $a \in A^*(n(\epsilon),\epsilon)$ and $0 \leq \epsilon \leq 1$.

(f) From lemma 2,

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

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

Now by (d),

$$| H_i [-\mathrm{tr}(H) + 2 H_i + 4 K_i] | \leq 4 (p+4) \|K(\epsilon)\|_\infty^2 \text{ for } i = 1,\dots,p.$$

Also, Stokes's theorem implies that

$$(-1)^{i-1} \int_{\partial A*} x_i^3 \, d\Phi^{(i)}(\mathbf{x}) = \int_{A*} (2x_i^2 + x_i^2 (1 - x_i^2)) \, d\Phi(\mathbf{x}) > 0$$

and

$$(-1)^{i-1} \int_{\partial A*} x_i x_j^2 \, d\Phi^{(i)}(\mathbf{x}) = \int_{A*} x_j^2 (1 - x_i^2) \, d\Phi(\mathbf{x}) > 0,$$

so

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

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

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

where $J(\mathbf{u})$ is the Jacobian

$$r^{-(p-1)} \frac{\partial [r \psi(u_1, u_2, \dots, u_{p-1})]}{\partial r \, \partial u_1 \, \partial u_2 \dots \partial u_{p-1}}.$$

Now $J(\mathbf{u})$ is always positive; hence for all $\mathbf{u}$ in Γ,

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

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

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

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

Thus $|\eta''(\epsilon)/\eta(\epsilon)| \leq (2p + 8 + 2(\tilde{\lambda}/\lambda) K) \|K(\epsilon)\|_\infty^2$.

(g) By an argument similar to that of (f),

$$|\eta'''(\epsilon)| \leq (\eta(\epsilon)/4) \{ 12 \|H\|_\infty^2 \|K\|_\infty [(\tilde{\lambda}/\lambda) K + 2]$$

$$+ 12\|\mathbf{H}\|_\infty |n''(\epsilon)/n(\epsilon)| [(\tilde{\lambda}/\Delta)K + 2] + 24\|\mathbf{K}\|_\infty [|n''(\epsilon)/n(\epsilon)| + \|\mathbf{H}\|_\infty\|\mathbf{K}\|_\infty]$$

$$+ \|\mathbf{H}\|_\infty{}^3 [(\tilde{\lambda}/\Delta)^2 K^2 + (4+2p)(\tilde{\lambda}/\Delta)K + 8 + 8p]\}.$$

Parts (d) and (f) then imply that

$$|n'''(\epsilon)| \leqq n(\epsilon)\|\mathbf{K}(\epsilon)\|_\infty{}^3 \{12[(\tilde{\lambda}/\Delta)K + 2]$$

$$+ 6[2(\tilde{\lambda}/\Delta)^2K^2 + (12+2p)(\tilde{\lambda}/\Delta)K + 4p + 16] + 6[2p + 10 + 2(\tilde{\lambda}/\Delta)K]$$

$$+ 2[(\tilde{\lambda}/\Delta)^2K^2 + (4+2p)(\tilde{\lambda}/\Delta)K + 8 + 8p]\}$$

$$\leqq n(\epsilon)\|\mathbf{K}(\epsilon)\|_\infty{}^3 (\tilde{\lambda}/\Delta)^2 K_2,$$

where $K_2 = 14K^2 + (104 + 16p)K + 196 + 52p$.

(h) The result follows from (4.5), part (a), and Theorem 5.6.7 of Graybill (1983). □

The partial derivatives of g with respect to n are then bounded in the following lemma.

Lemma 6. Suppose $n > 0$ and $0 < \epsilon < 1$. Then

(a) $0 < g_1(n,\epsilon) \leqq p/(2n)$.

(b) $|g_{11}(n,\epsilon)| \leqq [2n]^{-2}[p^2 + 2p + (n/n(\epsilon))^2(\tilde{\lambda}/\Delta)^2K^2]$.

(c) $|g_{111}(n,\epsilon)| \leqq (2n)^{-3}\, p\,[4p + (n/n(\epsilon))(\tilde{\lambda}/\Delta)K]^2$.

Proof.

(a) From Lemma 4, $\int_{A^*(n,\epsilon)} [p - \mathbf{x}^T\mathbf{x}]\, d\Phi(\mathbf{x}) > 0$, so $g_1(n,\epsilon) > 0$. The result

$$\int_{A^*(n,\epsilon)} [p - \mathbf{x}^T\mathbf{x}]\, d\Phi(\mathbf{x}) \leqq \int_{A^*(n,\epsilon)} p\, d\Phi(\mathbf{x}) \leqq p,$$

implies that $g_1(n,\epsilon) \leqq p/(2n)$.

(b) $|g_{11}(n,\epsilon)| = \left|(1/4n^2) \int_{A^*(n,\epsilon)} [[p - \mathbf{x}^T\mathbf{x}]^2 - 2p] \, d\Phi(\mathbf{x})\right|$

$$\leq (2n)^{-2} [p^2 + 2p + \int_{(n/\mathcal{n}(\epsilon))^{1/2} A^*(\mathcal{n}(\epsilon),\epsilon)} [\mathbf{x}^T\mathbf{x}]^2 \, d\Phi(\mathbf{x})].$$

Now by Lemma 5(e), $\mathbf{x}^T\mathbf{x} \leq (\tilde{\lambda}/\Delta)K$ on $A^*(\mathcal{n}(\epsilon),\epsilon)$. Hence

$$|g_{11}(n,\epsilon)| \leq (2n)^{-2} [2p + p^2 + (n/\mathcal{n}(\epsilon))^2 (\tilde{\lambda}/\Delta)^2 K^2].$$

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

$$|g_{111}(n,\epsilon)| \leq (2n)^{-3} \int_{A^*(n,\epsilon)} [\, p\,[p^2 + (\mathbf{x}^T\mathbf{x})^2] + 6p[p + \mathbf{x}^T\mathbf{x}] + 8p] \, d\Phi(\mathbf{x})]$$

$$\leq (2n)^{-3} [15p^3 + p\,(n/\mathcal{n}(\epsilon))^2 (\tilde{\lambda}/\Delta)^2 K^2 + 6p\,(n/\mathcal{n}(\epsilon))(\tilde{\lambda}/\Delta)K]. \quad \square$$

To find the expectations of the derivatives of $\mathcal{n}$ we work in the usual inner product space of pxp matrices, with

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

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

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

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

Lemma 7. Suppose that $n\,\text{vec}(\mathbf{W} - \mathbf{I}) \Rightarrow \mathcal{N}(\mathbf{0}, \mathbf{I} \otimes \mathbf{I} + \mathbf{C})$ for some n increasing to infinity and that the moments of $\mathbf{W}$ are bounded. Then

(a) $E[\lambda_p^{-j}] \leq p^j + o(1)$.

(b) $E[n^k \| \mathbf{W} - \mathbf{I} \|_\infty^k] = O(1)$.

(c) $E[\tilde{\lambda}^i \Delta^{-j} n^k \|\mathbf{W} - \mathbf{I}\|_\infty^k] = O(1)$.

(d) $E[(n\,\mathcal{n}'(0)/N)^2] = 2\,\text{tr}\,[\mathfrak{M}^2] + o(1)$.

(e) $E[n^2 \mathcal{n}''(0)/N] = [\int_{A'} [p - \mathbf{x}^T\mathbf{x}]\, d\Phi(\mathbf{x})]^{-1} [\int_{A'} \{2p\,(\mathrm{tr}\,\mathfrak{M}^2) - 4 + 2p$

$+ (p - \mathbf{x}^T\mathbf{x})[\mathbf{x}^T\mathbf{x}\,(\mathrm{tr}\,\mathfrak{M}^2 + 1) - 2\,\mathbf{x}^T\mathfrak{M}\mathbf{x}]\}\, d\Phi(\mathbf{x})]$

$+ (4 - p)\,\mathrm{tr}(\mathfrak{M}^2) - p - 2 + o(1)$

$= [\int_{A'} [p - \mathbf{x}^T\mathbf{x}]\, d\Phi(\mathbf{x})]^{-1} \sum (-1)^{i-1} \int_{\partial A^*} x_i$

$[\mathbf{x}^T\mathbf{x}\,(\mathrm{tr}\,\mathfrak{M}^2 + 1) - 2\,\mathbf{x}^T\mathfrak{M}\mathbf{x}]\, d\Phi^{(i)}(\mathbf{x})$

$- 4[\int_{A'} [p - \mathbf{x}^T\mathbf{x}]\, d\Phi(\mathbf{x})]^{-1} \int_{A'} [1 - \mathbf{x}^T\mathfrak{M}\mathbf{x}]\, d\Phi(\mathbf{x})$

$+ (6 - p)\,\mathrm{tr}(\mathfrak{M}^2) - p + o(1).$

Proof. Throughout the proof, let $\mathfrak{L} = \mathbf{W} - \mathbf{I}$ and let $\mathfrak{K} = (\mathcal{n}'(0)/N)\,\mathbf{I} - \mathfrak{L}$.

(a) Note that $\lambda^{-1} \leq \mathrm{tr}[\mathbf{W}^{-1}]$. Since $n\,\mathrm{vec}(\mathbf{W} - \mathbf{I}) \Rightarrow \mathcal{N}(\mathbf{0}, \mathbf{I} \otimes \mathbf{I} + \mathbf{C})$, $n\,(\mathrm{tr}[\mathbf{W}^{-1}] - p) \Rightarrow \mathcal{N}(0, 2p)$ by the delta method. Thus by dominated convergence, $E[\lambda^{-j}] \leq E[(\mathrm{tr}\,\mathbf{W}^{-1})^j] = p^j + o(1)$.

(b) For any even k, $E[\|\mathfrak{L}\|_\infty^k] \leq \sum E[\mathfrak{L}_{ij}^k]$. The entry $(n\,\mathfrak{L}_{ij})$ has either $\mathcal{N}(0,1)$ or $\mathcal{N}(0,2)$ as its limiting distribution. Thus for any even k, dominated convergence implies that

$$E[n^k \|\mathfrak{L}\|_\infty^k] \leq \sum E[n^k \mathfrak{L}_{ij}^k] \leq 2^k p^2 \alpha_k + o(1),$$

where α_k is the k^{th} moment of the standard normal distribution. The result for odd k follows from the Cauchy-Schwarz inequality.

(c) By Hölder's inequality,

$E[\tilde{\lambda}^i \lambda^{-j} n^k \|\mathfrak{L}\|_\infty^k]$

$$\leq \{E[\lambda^{-(j+1)}]\}^{j/(j+1)}\, \{E[(\tilde{\lambda}^i n^k \|\mathfrak{L}\|_\infty^k)^{j+1}]\}^{1/(j+1)}.$$

Now $\{E[\lambda^{-(j+1)}]\}^{j/(j+1)} \leq p^j + o(1)$ by part (a) of this lemma, and

$$E[\,(\,\tilde{\lambda}^i n^k \|\,\mathfrak{L}\,\|_\infty{}^k\,)^{j+1}\,] \leq E[\,(\,(\|\Lambda - \mathbf{I}\|_\infty + 1)^i\, n^k \|\,\mathfrak{L}\,\|_\infty{}^k\,]^{j+1}\,]$$

$$\leq E[\,(\,1 + \|\mathfrak{L}\|_\infty)^{i(j+1)}\, n^k \|\,\mathfrak{L}\|_\infty{}^{k(j+1)}\,] = O(1)$$

by (b).

(d) Changing variables, we rewrite $n'(0) = N\,\mathrm{tr}\,(\mathfrak{L}\mathfrak{M})$. The result follows since the asymptotic variance of $n\,\mathrm{tr}\,(\,\mathfrak{L}\mathfrak{M}\,) = n\,(\mathrm{vec}\,\mathfrak{M})^T(\mathrm{vec}\,\mathfrak{L})$ is

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

(e) Again changing variables and simplifying,

$$n''(0)/N = [2\textstyle\int_{A'} [p - \mathbf{x}^T\mathbf{x}]\, d\Phi(\mathbf{x})]^{-1}\, [2\gamma\mathrm{tr}(\mathfrak{K}^2) - \textstyle\int_{A'} [\mathrm{tr}(\mathfrak{K}) - \mathbf{x}^T\mathfrak{K}\mathbf{x}]^2 d\Phi(\mathbf{x})]$$

$$+ 2\,\mathrm{tr}(\mathfrak{K}\mathfrak{L}\mathfrak{M}). \tag{4.9}$$

We find the expectation of each term in (4.9) separately. The asymptotic variance of $(n\,\mathrm{vec}\,\mathfrak{K}) = n\,[(\mathrm{vec}\,\mathbf{I})(\mathrm{vec}\,\mathfrak{M})^T - \mathbf{I} \otimes \mathbf{I}]\,\mathrm{vec}\,\mathfrak{L}$ is

$$[\,(\mathrm{vec}\,\mathbf{I})(\mathrm{vec}\,\mathfrak{M})^T - \mathbf{I} \otimes \mathbf{I}\,]\,[\,\mathbf{I} \otimes \mathbf{I} + \mathbf{C}]\,[\,(\mathrm{vec}\,\mathbf{I})(\mathrm{vec}\,\mathfrak{M})^T - \mathbf{I} \otimes \mathbf{I}\,]^T$$

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

Thus, by dominated convergence,

$$E\,[\,n^2\,\mathrm{tr}(\mathfrak{K}^2)\,] = \mathrm{tr}\,E\,[\,n^2\,(\mathrm{vec}\,\mathfrak{K})(\mathrm{vec}\,\mathfrak{K})^T\,]$$

$$= 2\,p\,(\mathrm{tr}\,\mathfrak{M}^2) - 4 + p(p+1) + o(1). \tag{4.10}$$

Also,

$$E\,(\,n^2\,[\mathrm{tr}\,\mathfrak{K} - \mathbf{x}^T\mathfrak{K}\mathbf{x}]^2\,) = E\,(\,[n\,\mathrm{tr}\,(\mathfrak{K}\,(\mathbf{I} - \mathbf{x}\mathbf{x}^T))]^2\,)$$

$$= E\,[\,(\,n\,(\mathrm{vec}\,(\mathbf{I} - \mathbf{x}\mathbf{x}^T)\,)^T[\,(\mathrm{vec}\,\mathbf{I})(\mathrm{vec}\,\mathfrak{M})^T - \mathbf{I} \otimes \mathbf{I}\,]\,(\mathrm{vec}\,\mathfrak{L})\,)^2\,]$$

$$= 2\,(\mathrm{tr}\,[\mathbf{I} - \mathbf{x}\mathbf{x}^T])^2(\mathrm{tr}\,\mathfrak{M}^2\,) - 4\,(\mathrm{tr}\,[\mathbf{I} - \mathbf{x}\mathbf{x}^T])\,(\mathrm{tr}\,[\,[\mathbf{I} - \mathbf{x}\mathbf{x}^T]\,\mathfrak{M}])$$

$$+ 2\,\mathrm{tr}[\,(\mathbf{I} - \mathbf{x}\mathbf{x}^T)^2]$$

$$= 2(p - \mathbf{x}^T\mathbf{x})^2(\mathrm{tr}\,[\mathfrak{M}^2]\,) - 4\,(p - \mathbf{x}^T\mathbf{x})\,(1 - \mathbf{x}^T\mathfrak{M}\mathbf{x})$$

$$+ 2\,[p - 2\mathbf{x}^T\mathbf{x} + (\mathbf{x}^T\mathbf{x})^2]. \tag{4.11}$$

Finally part (b) implies that

$$E[\, n^2 \operatorname{tr}(\mathfrak{K}\mathfrak{L}\mathfrak{M}) \,] = E[\, \{n \operatorname{tr}(\mathfrak{M}\mathfrak{L})\}^2 - n^2 \operatorname{tr}(\mathfrak{M}\mathfrak{L}^2))]$$
$$= 2 \operatorname{tr}(\mathfrak{M}^2) - (p+1) + o(1). \tag{4.12}$$

The first expression of part (e) is thus proven by combining (4.9) through (4.12) and lemma 2, and the second expression results from applying Stokes's theorem to the first. □

Lemma 8. Let $H(x) = P\{ f(x,\mathbf{W}) \geqq \gamma \}$, and suppose that $n \operatorname{vec}(\mathbf{W} - \mathbf{I}) \Rightarrow \mathcal{N}(\mathbf{0}, \mathbf{I} \otimes \mathbf{I} + \mathbf{C})$ for some n increasing to infinity and that the moments of $\mathbf{W}$ are bounded. Then $E\,[\, [[n(1)]] - n(1)\,] = 1/2 + r$, where $|r| \le \int_0^\infty |H''(x)|\, dx + o(N^{-1/2})$.

Proof. Note that $P\{ n(1) \leqq x\} = P\{ f(x,\mathbf{W}) \geqq \gamma \} = H(x)$. From Hall (1981), the expectation of $R = [[n(1)]] - n(1)$ is

$$E\,[R] = 1/2 + \int r\, \{(1/2)\, H'(1-r) - H(1-r) + \sum \int H''(x-r)\,(x - n - 1/2)\, dx\}dr,$$

and the integral does not exceed $2H(1) + \int_0^\infty |H''(x)|\, dx$ in absolute value. But $H(1) \le P\{ \lambda_p \leqq N^{-1/2} \} \le N^{-1/2} E\,[\lambda_p^{-1}] = O(N^{-1/2})$ by lemma 7(a), completing the proof. □

Proof of Theorem 2.

Recall from (4.1) that

$$n(\epsilon) = N(\, \epsilon(1 + \ell/m)\, \hat{\boldsymbol{\Sigma}}_m + (1 - \epsilon)\mathbf{I})$$

so that the lemmas of this section may be applied with $\mathbf{W} = (1 + \ell/m)\, \hat{\boldsymbol{\Sigma}}_m$. From Wishart (1928), $E\,[\hat{\boldsymbol{\Sigma}}_m] = \mathbf{I}$ and $\operatorname{Cov}(\operatorname{vec} \hat{\boldsymbol{\Sigma}}_m) = (m-1)^{-1}\, (\mathbf{I}\otimes\mathbf{I}+\mathbf{C})$. In addition, all the moments of $\mathbf{W}$ are finite, so $\sqrt{m} \operatorname{vec}(\mathbf{W} - \mathbf{I}) \Rightarrow \mathcal{N}(\mathbf{0}, \mathbf{I}\otimes\mathbf{I}+\mathbf{C})$ by the multivariate central limit theorem and the dominated convergence theorem may be applied throughout the proof.

(a) The mean value theorem implies that $n(1) - N = n'(\epsilon)$ for some ϵ between 0 and 1, so by Lemma 5(c),

$$|\tau(\ell) - N| \leqq 1 + n(\epsilon)\|(1 + \ell/m)\hat{\Sigma}_m - \mathbf{I}\|_\infty \leqq 1 + (\tilde{\lambda}/\Delta)N\|(1 + \ell/m)\hat{\Sigma}_m - \mathbf{I}\|_\infty.$$

Now $(1 + \ell/m)\hat{\Sigma}_m$ converges almost surely to the identity matrix and $(\tilde{\lambda}/\Delta) \leqq (1 + \|(1 + \ell/m)\hat{\Sigma}_m - \mathbf{I}\|_\infty)(1 + \|\hat{\Sigma}_m^{-1} - \mathbf{I}\|_\infty)$, so $|\tau(\ell) - N|/N \to 0$ almost surely.

(b) By lemmas 5(b) and 7(c),

$$E[(\tau(\ell)/N)^q] \leq E[(1 + \tilde{\lambda})^q] < \infty \text{ for } q > 0$$

and

$$E[(\tau(\ell)/N)^q] \leq E[\Delta^q] \leq 1 + (1 + \ell/m)^q E[(1 + \|\hat{\Sigma}_m^{-1} - \mathbf{I}\|_\infty)^q] < \infty \text{ for } q < 0.$$

Hence by dominated convergence and part (a), $E[(\tau(\ell)/N)^q] \to 1$ as $N \to \infty$.

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

$$\tau(\ell) - N = (\tau(\ell) - n(1)) + n'(0) + (1/2)\,n''(0) + (1/6)\,n'''(\epsilon),$$

where ϵ is between 0 and 1. Now lemma 8 shows that

$$E[\tau(\ell) - n(1)] = 1/2 + \int_0^\infty |H''(x)|\,dx + o(N^{-1/2}).$$

Here $H(x) = P\{\hat{\Sigma}_m/x \in (1 + \ell/m)^{-1} R_\gamma(A)\}$, where $R_\gamma(A) = \{\mathbf{V}: \Phi(\mathbf{V}^{-1/2}A) \geq \gamma\}$. Direct computation using the Wishart density then shows that $\int_0^\infty |H''(x)|\,dx = o(1)$. Also, lemmas 5 and 7 imply that

$$E[m^{3/2}|n'''(\epsilon)|] \leqq m^{3/2} N K_2 E[\tilde{\lambda}^3 \Delta^{-5}\|(1 + \ell/m)\hat{\Sigma}_m - \mathbf{I}\|_\infty^3] = O(1),$$ so $E[|n'''(\epsilon)|] = o(N/m)$. Thus $E[\tau(\ell) - N] = 1/2 + E[n'(0)] + (1/2)E[n''(0)] + o(N/m)$, where $E[n'(0)] = E[\operatorname{tr}\{[(1 + \ell/m)\hat{\Sigma}_m - \mathbf{I}]\mathfrak{M}\}] = \ell/m$ and $E[n''(0)]$ is given

explicitly in lemma 7(e).

(d) We square the second-order Taylor expansion of $n(1)$ about 0 to give

$$E[(\tau(\ell)-N)^2] = E[\{n'(0)\}^2 + (1/4)\{n''(\epsilon)\}^2 + n'(0)\, n''(\epsilon)].$$

$E[\{n'(0)\}^2]$ is shown to equal $2N^2 \operatorname{tr}[\mathfrak{M}^2]/m + o(N/m)$ in lemma 7, and the other terms are shown to be $o(N/m)$ by applying lemmas 5 and 7.

(e) A second-order Taylor series expansion gives

$$\sqrt{m}\,(\tau(\ell)-N)/N = \sqrt{m}\,[\tau(\ell) - n(1) + n'(0) + (1/2)\, n''(\epsilon)]/N$$

for some ϵ between 0 and 1. Now $E[|n''(\epsilon)|] = o(1)$ by lemmas 5 and 7(c), so $\sqrt{m}\,|n''(\epsilon)|/N$ converges in probability to zero. Thus the limiting distribution of $\sqrt{m}\,(\tau(\ell)-N)/N$ is the same as that of

$\sqrt{m}\, n'(0)/N = \sqrt{m}\operatorname{tr}[((1+\ell/m)\hat{\boldsymbol{\Sigma}}_m - \mathbf{I})\mathfrak{M}]$, shown to be $\mathcal{N}(0, 2\operatorname{tr}[\mathfrak{M}^2])$ in the proof of Lemma 7(d).

(f) Since $f(n,\boldsymbol{\Sigma}) = g(n,0)$, (2.7) may be rewritten as

$$\begin{aligned} E[g(\tau(\ell),0)] &= \gamma + g_1(N,0)\, E[\tau(\ell)-N] + (1/2)\, g_{11}(N,0)\, E[(\tau(\ell)-N)^2] \\ &\quad + (1/2)\, E[\{g_{11}(n^*,0) - g_{11}(N,0)\}\,(\tau(\ell)-N)^2], \end{aligned}$$

where n^* is between $\tau(\ell)$ and N. Using results (c) and (d), then,

$$\begin{aligned} E[g(\tau(\ell),0)] &= \gamma + g_1(N,0)\{N\ell/m + E[n''(0)]\} + (1/2)\, g_{11}(N,0)\, E[(n'(0))^2] \\ &\quad + (1/2)\, E[\{g_{11}(n^*,0) - g_{11}(N,0)\}\,(\tau(\ell)-N)^2] \\ &\quad + E[g_1(N,0)\, o(N/m) + g_{11}(N,0)\, o(N^2/m)], \end{aligned}$$

where $g_1(N,0) = (1/2N)\int_{A'} [p - \mathbf{x}^T\mathbf{x}]\, d\Phi(\mathbf{x})$ and

$g_{11}(N,0) = (1/4N^2)\int_{A'} \{[p - \mathbf{x}^T\mathbf{x}]^2 - 2p\}\, d\Phi(\mathbf{x})$. The inequalities in lemmas 5 and 6 imply that $E[g_1(N,0)\, o(N/m) + g_{11}(N,0)\, o(N^2/m)] = o(m^{-1})$.

The proof is completed by showing that

$$E[\,|g_{11}(n^*,0) - g_{11}(N,0)|\,(\tau(\ell) - N)^2]$$

is also $o(m^{-1})$. By the mean value theorem,

$E[\,|g_{11}(n^*,0) - g_{11}(N,0)|\,(\tau(\ell) - N)^2] = E[\,|g_{111}(n',0)|\,(\tau(\ell) - N)^2]$ for some n' between n^* and N. Since $n(\epsilon)$ is a continuously differentiable function of ϵ, $n' = n(\epsilon')$ for some ϵ' between 0 and 1. Lemmas 5 and 6 then demonstrate that

$$|g_{111}(n(\epsilon^*),0)| \leq (2\,n(\epsilon^*))^{-3}\,p\,[4p + (\,n(\epsilon^*)/N\,)(\tilde{\lambda}/\Delta)\,K]^2$$
$$\leq p\,(8N)^{-3}\,(4p + K)^2\,(\tilde{\lambda}^3/\Delta^5).$$

Thus

$$E[\,|g_{111}(n',0)|\,(\tau(\ell) - N)^2] \leq E[\,p\,(8N)^{-3}\,(4p + K)^2\,(\tilde{\lambda}^3/\Delta^5)\,(\tau(\ell) - N)^2] = o(m^{-1})$$

by lemma 7(c), part (b), and the Cauchy-Schwarz inequality. This completes the proof of the theorem. □

Proof of Corollary 1. Using the mean value theorem,

$$N[(1 + \hat{\ell}_2/m)\hat{\Sigma}_m] - N[(1 + \ell_2/m)\hat{\Sigma}_m]$$
$$= N[(1 + \ell_2^*/m)\hat{\Sigma}_m]\,[(1 + \ell_2^*/m)\,m]^{-1}\,(\hat{\ell}_2 - \ell_2)$$

for some ℓ_2^* between $\hat{\ell}_2$ and ℓ_2. Applying lemma 5, then,

$$|\,N[(1 + \hat{\ell}_2/m)\hat{\Sigma}_m] - N[(1 + \ell_2/m)\hat{\Sigma}_m]\,|$$
$$\leq (N/m)\,\|\,(1 + (\hat{\ell}_2 + \ell_2)/m)\hat{\Sigma}_m\|_\infty\,(\hat{\ell}_2 - \ell_2). \qquad (4.13)$$

We show that $(\hat{\ell}_2 - \ell_2)$ is small except on a set of small probability. Let $\delta = m^{-1/2}$ and $B = \{\,\|\hat{\Sigma}_m - \mathbf{I}\|_\infty \leq \delta\,\}$. On the set B, the symmetric difference of the sets $\hat{A}'$ and A' tends to the empty set: using Lemma 5 and the relationships between different matrix norms,

$$\|\hat{\Sigma}_m^{-1/2} - \mathbf{I}\|_\infty I_B \leq \sqrt{p}\,\|\Lambda^{-1/2} - \mathbf{I}\|_\infty I_B \leq \sqrt{p}\;\delta,$$

where Λ is the matrix of eigenvalues of $\hat{\Sigma}_m$. Thus

$$(\hat{A}' - A')\, I_B \subset \|(\tau(0)/N)^{1/2}\hat{\Sigma}_m^{1/2} - \mathbf{I}\|_\infty A' \subset 2\sqrt{p}\;\delta A',$$

so by lemma 5(e),

$$\|\int_{\hat{A}'} \mathbf{x}\mathbf{x}^T d\Phi(\mathbf{x}) - \int_{A'} \mathbf{x}\mathbf{x}^T d\Phi(\mathbf{x})\|_\infty I_B \leq \int_{2p\delta A'} \|\mathbf{x}\mathbf{x}^T\|_\infty d\Phi(\mathbf{x}) \leq 2p\,\delta K.$$

This result then implies that

$$|\ell_2 - \hat{\ell}_2|\, I_B \leq r(\delta), \tag{4.14}$$

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

We then show that $|\ell_2 + \hat{\ell}_2|$ is bounded by a function of Λ on B^c. Equation (2.6) implies that $\mathrm{tr}(\mathfrak{M}^2) \leq 1$ and $\mathrm{tr}(\hat{\mathfrak{M}}^2) \leq 1$. We use results from Lemma 7 to bound the remaining terms in $|\ell_2 + \hat{\ell}_2|$. By part (e) of that lemma and equation (2.6),

$$-(1/2)\,[\int_{A'} [p - \mathbf{x}^T\mathbf{x}]\,d\Phi(\mathbf{x})]^{-1} \int_{A'} (2p - 4 + (p - \mathbf{x}^T\mathbf{x})[\mathbf{x}^T\mathbf{x} - 2\,\mathbf{x}^T\mathfrak{M}\mathbf{x}])\,d\Phi(\mathbf{x})$$

$$\leq [\int_{A'} [p - \mathbf{x}^T\mathbf{x}]\,d\Phi(\mathbf{x})]^{-1} \int_{A'} (\,2 + (p - \mathbf{x}^T\mathbf{x})\,\mathbf{x}^T\mathfrak{M}\mathbf{x}\,)\,d\Phi(\mathbf{x})$$

$$= [\int_{A'} [p - \mathbf{x}^T\mathbf{x}]\,d\Phi(\mathbf{x})]^{-1}\,[\sum (-1)^{i-1} \int_{\partial A'} x_i\,\mathbf{x}^T\mathfrak{M}\mathbf{x}\,d\Phi^{(i)}(\mathbf{x})$$

$$+ 2\int_{A'} (1 - \mathbf{x}^T\mathfrak{M}\mathbf{x})\,d\Phi(\mathbf{x})]$$

$$\leq \|\mathfrak{M}\|_\infty + 2\,\mathrm{tr}(\mathfrak{M}^2)$$

$$\leq pK + 2.$$

Similarly, using Lemma 5(e),

$$-(1/2)\,[\int_{\hat{A}'} [p - \mathbf{x}^T\mathbf{x}]\,d\Phi(\mathbf{x})]^{-1} \int_{\hat{A}'} (2p - 4 + (p - \mathbf{x}^T\mathbf{x})[\mathbf{x}^T\mathbf{x} - 2\,\mathbf{x}^T\hat{\mathfrak{M}}\mathbf{x}])\,d\Phi(\mathbf{x})$$

$$\leq p\,K\,\lambda_1\,\lambda_p^{-1} + 2.$$

Consequently,

$$(\hat{\ell}_2 + \ell_2) \leq p + 6 + 2\,p\,K\,\lambda_1\,\lambda_p^{-1}. \tag{4.15}$$

Equations (4.13), (4.14), and (4.15) imply that

$$\begin{aligned}
&\left| N[(1 + \hat{\ell}_2/m)\hat{\Sigma}_m] - N[(1 + \ell_2/m)\hat{\Sigma}_m]\right| \\
&\qquad \leq m^{-1}N\,\lambda_1\,(1 + (\ell_2 + |\,\hat{\ell}_2 - \ell_2|)/m)\,|\,\hat{\ell}_2 - \ell_2| \\
&\qquad \leq m^{-1}N\,\lambda_1\,(\,r(\delta) + O(m^{-1})) \\
&\qquad\qquad + m^{-1}N\,\lambda_1\,K_6\,(\tilde{\lambda}/\Delta)\,(2 + K_6\,(\tilde{\lambda}/\Delta)/m)\; I_{B^c}.
\end{aligned} \tag{4.16}$$

Inequality (4.16) and the fact that $P(B^c) = o(1)$ by Chebychev's inequality are then used to prove the corollary. Since λ_1 and $(\tilde{\lambda}/\Delta)$ tend to 1 almost surely as N tends to infinity, and since $I_{B^c} \to 0$ almost surely, (4.16) implies that $|\,\tau(\hat{\ell}_2) - \tau(\ell_2)\,|\,/N \to 0$ almost surely as $N \to \infty$. This proves part (a) of the corollary. Part (b) follows since the q^{th} power of the right-hand side of (4.16) is dominated by a function with finite expectation, as in the proof of theorem 2.

To prove (c), note that

$$\begin{aligned}
E[\,\tau(\hat{\ell}_2) - \tau(\ell_2)\,] &= E\,[\,\tau(\hat{\ell}_2) - N[(1 + \hat{\ell}_2/m)\hat{\Sigma}_m]] \\
&\quad - E[\,\tau(\ell_2) - N[(1 + \ell_2/m)\hat{\Sigma}_m]] + E\,[\,N[(1 + \hat{\ell}_2/m)\hat{\Sigma}_m] - N[(1 + \ell_2/m)\hat{\Sigma}_m]\,].
\end{aligned}$$

The first two expectations on the right-hand-side are both equal to 1/2 + o(1) by lemma 8 and the proof of Theorem 2(c). The Cauchy-Schwarz inequality and inequality (4.16) imply that

$$E\,[\,\left|\,N[(1 + \hat{\ell}_2/m)\hat{\Sigma}_m] - N[(1 + \ell_2/m)\hat{\Sigma}_m]\right|$$

$$\leq m^{-1} N\, E[\, \lambda_1(\, r(\delta) + K_6(\tilde{\lambda}/\Delta)(2 + K_6(\tilde{\lambda}/\Delta)/m)\, I_{B^c}]$$

$$= o(N/m), \tag{4.17}$$

proving (c).

It is similarly shown that $E[\, (\, \tau(\hat{\ell}_2) - \tau(\ell_2)\,)^2\,] = o(N^2/m^2)$, proving part (d) of the corollary. Inequality (4.17) also implies that

$$E\,[\sqrt{m}\, |\, \tau(\hat{\ell}_2) - \tau(\ell_2)\,|\, /N\,] \to 0 \text{ as } N \to \infty.$$

Thus $\sqrt{m}\,|\tau(\hat{\ell}_2) - \tau(\ell_2)|\,/N$ converges in probability to zero, so $\sqrt{m}\ \tau(\hat{\ell}_2)/N$ has the same limiting distribution as $\sqrt{m}\ \tau(\ell_2)/N$. This proves part (e).

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

$$E[g(\tau(\hat{\ell}_2),0)] = E[g(\tau(\ell_2),0)] + E[\, g_1(\tau(\ell_2{}^*),0)\,(\tau(\hat{\ell}_2) - \tau(\ell_2))],$$

where $\ell_2{}^*$ is between ℓ_2 and $\hat{\ell}_2$. Lemma 6(a), the result from the proof of part (d) that $E[\, (\, \tau(\hat{\ell}_2) - \tau(\ell_2)\,)^2\,] = o(N^2/m^2)$ and the Cauchy-Schwarz inequality imply that

$$E[g_1(\tau(\ell_2{}^*),0)\,|\tau(\hat{\ell}_2) - \tau(\ell_2)|\,] \leq \{\, E\,[\,(p/\tau(\ell_2{}^*))^2\,]\, E[\,(\tau(\hat{\ell}_2) - \tau(\ell_2))^2]\,\}^{1/2}$$
$$= o(m^{-1})$$

by part (b) of this corollary. Thus

$$E[g(\tau(\hat{\ell}_2),0)] = E[g(\tau(\ell_2),0)] + o(m^{-1}) = \gamma + o(m^{-1})$$

by Theorem 2(f) and the definition of ℓ_2 in (2.4). □

The proof of theorem 3 and its corollary resembles that of theorem 2, with the added complication that $\hat{\Sigma}_{t_2}$ no longer follows a Wishart distribution, but in fact systematically underestimates Σ.

Let $\overline{X}_{(1)} = \overline{X}_m$, $\hat{\Sigma}_{(1)} = \hat{\Sigma}_m$, and let $\overline{X}_{(2)}$ and $\hat{\Sigma}_{(2)}$ be the least squares estimates of the mean and covariance using only the observations in the second sample. Then the estimated covariance using both samples may be expressed as a function of $\overline{X}_{(1)}$, $\hat{\Sigma}_{(1)}$, $\overline{X}_{(2)}$, $\hat{\Sigma}_{(2)}$, and t_2, as is shown in the following lemma.

Lemma 9.

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

Lemma 9 is used in the following lemma to evaluate the conditional expectation of $\hat{\Sigma}_{t_2}$.

Lemma 10.

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

Proof. The result follows from Lemma 9 because

$$[\overline{X}_{(1)} - \overline{X}_{(2)}] \mid \hat{\Sigma}_m \sim \mathcal{N}(\mathbf{0}, [m^{-1} + (t_2 - m)^{-1}]\mathbf{I}). \quad \square$$

We then may approximate the moments of t_2 and $\hat{\Sigma}_{t_2}$.

Lemma 11. Suppose the conditions of theorem 4 hold. Then

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

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

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

Proof. Let

$$D \equiv \{\tau(0) > m\}$$

and let

$$C \equiv \{ \| \hat{\Sigma}_m - \mathbf{I} \|_\infty \leqq 1/2 \}.$$

Lemma 5 implies that $| \tau(0) - N | \, I_C \leqq N/2$. Then, using Cramér's theorem on large deviations (see Varadhan (1984)),

$$P\{C^c\} \leqq 2\,p^2 \exp[\,-(m-1)/24]$$

and hence

$$P\{D^c\} \leqq P\{\,\{ |\tau(0) - N| > N-m \} \cap C\} + P\{C^c\} \leqq 2\,p^2 \exp[\,-(m-1)/24] \qquad (4.18)$$

for sufficiently large N. Thus

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

and

$$E[\,|t_2 - N_2|^{\,j} / t_2^{\,k}\, I_D] = E[\,| \,[[c(\tau(0)-m)]] - c(N-m)\,|^j \,/ (\,[[c(\tau(0)-m)]] + m)^k \, I_D]$$

$$\leqq m^{-k} + c\,E[\,|\,\tau(0) - N\,|^{\,j} \,/\, \tau(0)^{\,k}] = o(N^{j-k})$$

by theorem 2(b), completing the proof of (a).

To prove part (b), note that

$$E[t_2^{-k}] = N_2^{-k}\, E[(1 - (t_2 - N_2)/t_2)^k]$$

$$= N_2^{-k} + N_2^{-k} \sum_{j=1}^{k} \binom{k}{j} E[((t_2 - N_2)/t_2)^j]$$

by the Binomial theorem. The result then follows from (a).

For part (c), Lemma 10 implies that

$$E[\,\hat{\Sigma}_{t_2}\,] = E\,\{\,E[\,\hat{\Sigma}_{t_2} | \,\hat{\Sigma}_m\,]\,\} = \mathbf{I} \; + E[\,(t_2 - 1)^{-1}(m-1)\,(\hat{\Sigma}_m - \mathbf{I})\,].$$

Now

$$(t_2-1)^{-1} = N_2^{-1} - N_2^{-2}\,(t_2-N_2-1) + N_2^{-2}\,(t_2-1)^{-1}\,(t_2-N_2-1)^2 \quad .$$

so

$$E[(t_2-1)^{-1}(m-1)(\hat{\Sigma}_m - \mathbf{I})] = (m-1)N_2^{-2}[-c\,E[(\hat{\Sigma}_m - \mathbf{I})(\tau(0)-N)]$$
$$+ E[(\hat{\Sigma}_m - \mathbf{I})\{[[c(\tau(0)-m)]] - c(\tau(0)-m)\}\, I_D]$$
$$- c\,E[(\hat{\Sigma}_m - \mathbf{I})(\tau(0)-m)\, I_{D^c}]$$
$$+ E[(\hat{\Sigma}_m - \mathbf{I})(t_2-1)^{-1}(t_2-N_2-1)^2]]. \quad (4.19)$$

We find the expectations of each term in (4.19) separately. A second-order Taylor series expansion gives that for some ϵ^* between 0 and 1,

$$E[(\hat{\Sigma}_m - \mathbf{I})(\tau(0)-N)] = E[(\hat{\Sigma}_m - \mathbf{I})\{n'(0) + (1/2)\,n''(\epsilon^*)\}]$$
$$= N\,E[(\hat{\Sigma}_m - \mathbf{I})\,\mathrm{tr}\,\{\mathfrak{M}(\hat{\Sigma}_m - \mathbf{I})\}] + (1/2)\,E[(\hat{\Sigma}_m - \mathbf{I})\,n''(\epsilon^*)].$$

Now $E[(\hat{\Sigma}_m - \mathbf{I})\,\mathrm{tr}\,\{\mathfrak{M}(\hat{\Sigma}_m - \mathbf{I})\}] = 2(m-1)^{-1}\mathfrak{M}$, it was shown in the proof of theorem 2(d) that $E[\,|n''(\epsilon^*)|^2\,] = o(N^2/m)$, and lemma 7(b) implies that $E[\,\|\hat{\Sigma}_m - \mathbf{I}\|_\infty^2\,] = O(m^{-1})$, so $E[\,\|\hat{\Sigma}_m - \mathbf{I}\|_\infty\,|n''(\epsilon^*)|\,] = o(N/m)$ by the Cauchy-Schwarz inequality. Combining terms,

$$E[(\hat{\Sigma}_m - \mathbf{I})(\tau(0)-N)] = 2N(m-1)^{-1}\mathfrak{M} + o(N/m). \quad (4.20)$$

It is easily seen that

$$E[\,\|\hat{\Sigma}_m - \mathbf{I}\|_\infty\,|[[c(\tau(0)-m)]] - c(\tau(0)-m)|\, I_D] \le E[\,\|\hat{\Sigma}_m - \mathbf{I}\|_\infty\,] = o(1). \quad (4.21)$$

Also, equation (4.18) and the Cauchy-Schwarz inequality imply that

$$E[\,\|\hat{\Sigma}_m - \mathbf{I}\|_\infty\,|\tau(0)-m|\, I_{D^c}] \le m\,\{\mathrm{tr}\,E[(\hat{\Sigma}_m-\mathbf{I})^2]\,P(D^c)\}^{1/2}$$
$$\le 2mp\,(p+1)\exp[-(m-1)/48] = o(1). \quad (4.22)$$

To show that $E[(\hat{\Sigma}_m - \mathbf{I})(t_2-1)^{-1}(t_2-N_2-1)^2]$ is $o(N/m)$, let $\delta = m^{-1/2}$ and let

$$B \equiv \{\,\|\hat{\Sigma}_m - \mathbf{I}\|_\infty \le \delta\,\} \cap D.$$

By Cramér's theorem (Varadhan (1984)), $P(B^c) \leq (2p+p^2) \exp\{-m^{1/4}/12\}$, so $E[\, \|\hat{\boldsymbol{\Sigma}}_m - \mathbf{I}\|_\infty (t_2-1)^{-1} (t_2-N_2-1)^2 I_{B^c}\,] = o(N/m)$. Then, using lemma 5(b) to show that $|\tau(0)-N|\, I_B \leq \delta N$,

$$E[\, \|\hat{\boldsymbol{\Sigma}}_m - \mathbf{I}\|_\infty (t_2-1)^{-1} (t_2-N_2-1)^2 I_B\,] \leq \delta^3 N_2/(1-\delta) = o(N/m).$$

Thus $E[\, \|\hat{\boldsymbol{\Sigma}}_m - \mathbf{I}\|_\infty (t_2-1)^{-1} (t_2-N_2-1)^2\,] = o(N/m)$. (4.23)

Thus, using (4.19) through (4.23),

$$E[\,(m-1)(t_2-1)^{-1}(\hat{\boldsymbol{\Sigma}}_m - \mathbf{I})\,] = -2\,\mathfrak{M}/N_2 + o(N^{-1}),$$

completing the proof of (c). □

Proof of Theorem 3. For any constant n, Graybill (1983, Theorem 10.10.1) implies that $E[\hat{\boldsymbol{\Sigma}}_n] = \mathbf{I}$ and $\mathrm{Cov}(\mathrm{vec}\, \hat{\boldsymbol{\Sigma}}_n) = (n-1)^{-1}(\mathbf{I} \otimes \mathbf{I} + \mathbf{C})$. By the multivariate central limit theorem, then,

$$\sqrt{n}\, \mathrm{vec}[\hat{\boldsymbol{\Sigma}}_n - \mathbf{I}] \Rightarrow \mathcal{N}(\mathbf{0}, \mathbf{I} \otimes \mathbf{I} + \mathbf{C}).$$

Now for any constant n, $\sqrt{n}\, \mathrm{vec}[\hat{\boldsymbol{\Sigma}}_n - \mathbf{I}]$ is uniformly continuous in probability since it may be rewritten as a normalized partial sum. Since (t_2/N_2) converges in probability to one by Lemma 11(b), and since ℓ/t_2 converges in probability to zero, Anscombe's (1952) theorem implies that

$$\sqrt{N_2}\, \mathrm{vec}[(1+\ell/t_2)\hat{\boldsymbol{\Sigma}}_{t_2} - \mathbf{I}] \Rightarrow \mathcal{N}(\mathbf{0}, \mathbf{I} \otimes \mathbf{I} + \mathbf{C}).$$

Dominated convergence may be applied throughout the proof since by lemma 9, for $k \geq 0$, $E[\hat{\boldsymbol{\Sigma}}_{t_2}^{k}] \leq E[(\hat{\boldsymbol{\Sigma}}_{(1)} + \hat{\boldsymbol{\Sigma}}_{(2)} + [\mathbf{X}_{(1)} - \mathbf{X}_{(2)}][\mathbf{X}_{(1)} - \mathbf{X}_{(2)}]^T)^k]$, and the expectation on the right is shown to be finite by using successive conditioning. Recall from (4.1) that for triple sampling,

$$n(\epsilon) = N(\ \epsilon(1 + \ell/t_2)\ \hat{\Sigma}_{t_2} + (1 - \epsilon)\mathbf{I}) \qquad (4.24)$$

so that lemmas 1 through 8 may be applied with $\mathbf{W} = (1 + \ell/t_2)\ \hat{\Sigma}_{t_2}$.

Results (a) and (b) follow the same proofs as parts (a) and (b) of theorem 2 once it is noted that $(1 + \ell/t_2)\hat{\Sigma}_{t_2}$ converges almost surely to the identity matrix since $\hat{\Sigma}_m \to \mathbf{I}$ almost surely and since t_2 is defined to be larger than m.

(c) As in the proof of theorem 2(c),

$$E[\ t_3(\ell) - N] = E[\ t_3(\ell) - n(1)] + E[\ n'(0)] + (1/2)E[n''(0)] + (1/6)\,E[\ n'''(\epsilon)], \qquad (4.25)$$

where ϵ is between 0 and 1. Lemma 5(g,h) and lemma 7(c) imply that $E\,[\,|\,n'''(\epsilon)|\,] = o(1)$, and $E\,[n''(0)/c]$ is evaluated explicitly in lemma 7(e), so the proof of (c) is completed by evaluating $E[\ t_3(\ell) - n(1)]$ and $E[\ n'(0)]$.

$$\begin{aligned} E\,[\ n'(0)\ ] &= N\ \mathrm{tr}\ [\ \mathfrak{M}\, E\ \{\ (1 + \ell/t_2)\ \hat{\Sigma}_{t_2} - \mathbf{I}\ \}\ ] \\ &= N\ \mathrm{tr}\ [\ \mathfrak{M}\, E\ \{\ (1 + \ell/N_2)\ \hat{\Sigma}_{t_2}\ \}\ ] \\ &\qquad - N\ \ell N_2^{-1}\ \mathrm{tr}\ [\ \mathfrak{M}\, E\ \{\ (t_2 - N_2)\, t_2^{-1}\ \hat{\Sigma}_{t_2}\ \}\ ]. \end{aligned}$$

Now $E\ \{\ \hat{\Sigma}_{t_2}\ \} = \mathbf{I} - 2\ \mathfrak{M}/N_2 + o(N^{-1})$ by Lemma 11(c). Also,

$$E\ \{\ (t_2 - N_2)\, t_2^{-1}\ \|\ \hat{\Sigma}_{t_2}\|_\infty\ \} \le \{\ E[(t_2 - N_2)^2 t_2^{-2}]\ E\ [\ \|\ \hat{\Sigma}_{t_2}\|_\infty^{\ 2}\ ]^{1/2} = o(1)$$

by parts (a) and (b) of Lemma 11 and the Cauchy-Schwarz inequality. Thus, since the entries of $\mathfrak{M}$ are bounded,

$$\begin{aligned} E\ [\ n'(0)\ ] &= N\ \mathrm{tr}\ [\ \mathfrak{M}\ \{\ (\ \ell/N_2)\mathbf{I} - 2\ \mathfrak{M}/N_2\ \} + o(1) \\ &= \ell/c - 2\ [\ \mathrm{tr}\ \mathfrak{M}^2]/c + o(1). \end{aligned}$$

From lemma 8, $E\ [\ (\ t_3(\ell) - n(1))\ ] = 1/2 + \int_0^\infty |H''(x)|\,dx\ + o(1)$. To show that $\int_0^\infty |H''(x)|\, dx = o(1)$, note that by using a Helmert-type transformation, $\hat{\Sigma}_n = (n-1)^{-1} \sum \mathbf{Y}_i \mathbf{Y}_i^T$, where $\mathbf{Y}_i = \{i(i+1)\}^{-1/2}(i\mathbf{X}_{i+1} - \sum \mathbf{X}_j)$, so that the $\mathbf{Y}_i$ are independent $\mathcal{N}(\mathbf{0},\mathbf{I})$ random vectors. Let $R = \{\mathbf{V}: \Phi(\mathbf{V}^{-1/2}A) \geq \gamma\}$. Then $H(x) = P\{ \hat{\Sigma}_{t_2}/x \in R\} = E[P\{ (t_2-1)^{-1} (\sum \mathbf{Y}_i \mathbf{Y}_i^T + (m-1)\hat{\Sigma}_m)/x \in R \mid \hat{\Sigma}_m \}]$, 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 $E[(t_3(\ell) - n(1))] = 1/2 + o(1)$.

(d) We square the right-hand side in (4.25) to give

$$E[(t_3(\ell) - N)^2] = E[\{ n'(0)\}^2] + \text{remainder terms;}$$

$E[\{ n'(0)\}^2]$ is given explicitly in lemma 7(d), and the remaining terms are shown to be o(N) by lemma 5.

(e) By equation (4.25),

$$\sqrt{N_2}\, (t_3(\ell) - N)/N = \sqrt{N_2}\, [n'(0) + (1/2)\, n''(\epsilon)] / N$$

for some ϵ between 0 and 1. Now $\sqrt{N_2}\, n'(0)/N = \sqrt{N_2}\, \mathrm{tr}\, [(1+ \ell/t_2)\, \hat{\Sigma}_{t_2} \mathfrak{M}]$ converges in distribution to a $\mathcal{N}(0,2\, \mathrm{tr}\, [\mathfrak{M}^2])$ random variable by Anscombe's (1952) theorem, and $E[| n''(\epsilon) |] = o(1)$ by lemmas 5(f) and 7, so the limiting distribution of $\sqrt{N_2}\, (t_3(\ell) - N)/N$ is $\mathcal{N}(0,2\, \mathrm{tr}\, [\mathfrak{M}^2])$.

(f) The proof of (f) depends on the following second-order Taylor expansion of g about its first argument,

$$E[g(t_3(\ell),0)] = \gamma + g_1(N,0)\, E[t_3(\ell) - N] + (1/2)\, g_{11}(N,0)\, E[(t_3(\ell) - N)^2] + (1/2)\, E[\{g_{11}(n^*,0) - g_{11}(N,0)\}\, (t_3(\ell) - N)^2],$$

where n^* is between $t_3(\ell)$ and N. Using results (c) and (d), then,

$$E[g(t_3(\ell),0)] = \gamma + g_1(N,0)\{\ell/c - 2(\mathrm{tr}\,\mathfrak{M}^2)/c + E[n''(0)]\}$$
$$+ (1/2)\, g_{11}(N,0)\, E[(n'(0))^2]$$
$$+ (1/2)\, E[\{g_{11}(n^*,0) - g_{11}(N,0)\}(t_3(\ell) - N)^2]$$
$$+ o(N^{-1})$$

since $E[\,|g_1(N,0)|\,] \leq p/2N$ and $E[\,|g_{11}(N,0)|\,] \leq (2N)^{-2}\, 9(p^2 + 2p + K^2)$ by lemma 6. The proof is completed by showing that

$E[\,|g_{11}(n^*,0) - g_{11}(N,0)|\,(t_3(\ell) - N)^2]$ is also $o(N^{-1})$. By the mean value theorem and lemma 6(b),

$$E[|g_{11}(n^*,0) - g_{11}(N,0)|\,(t_3(\ell) - N)^2] \leq E[\sup_{\epsilon^*} |g_{111}(n(\epsilon^*),0)|\,(t_3(\ell) - N)^2]$$
$$\leq E[\,\sup_{\epsilon^*} (2\,n(\epsilon^*))^{-3}\, p[4p + (n(\epsilon^*)/N)(\tilde{\lambda}/\Delta)K]^2\,(t_3(\ell) - N)^2],$$

where the supremum is taken over all ϵ^* between 0 and 1. Thus

$$E[\,|g_{11}(n^*,0) - g_{11}(N,0)|\,(t_3(\ell) - N)^2]$$
$$\leq E[(2p^3+K^2+1)(n(\epsilon^*)^{-3} + n(\epsilon^*)^{-2}(\tilde{\lambda}/\Delta) + n(\epsilon^*)^{-1}(\tilde{\lambda}/\Delta)^2)(t_3(\ell) - N)^2]$$
$$\leq 3N^{-3}(2p^3+K^2+1)\, E[\,(\tilde{\lambda}^3/\Delta^5)\,\|(1 + \ell/t_2)\,\hat{\Sigma}_{t_2} - \mathbf{I}\,\|^2] = o(N^{-1})$$

by lemmas 5(c) and 7(c). This completes the proof of the theorem. □

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. (1987). A theory of accurate estimation for multivariate data with applications to large-scale Monte Carlo studies. To appear.

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.* Belmont, CA: Wadsworth.

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.* New York: Marcel Dekker.

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

Lohr, S. (1987) *Accurate Multivariate Estimation using Double and Triple Sampling.* Unpublished Ph.D. dissertation, University of Wisconsin-Madison.

Miller, R.G. (1981). *Simultaneous Statistical Inference,* 2nd ed. New York: Springer-Verlag.

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

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.* New York: W.A. Benjamin, Inc.

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, Office of Solid Waste and Emergency Response. (1982). *Test methods for evaluating solid waste.* U.S. EPA Publication SW-846.

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

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.