Ann. Inst. Statist. Math.
Vol. 51, No. 4, 621–636 (1999)

ON A TWO-STAGE PROCEDURE HAVING SECOND-ORDER PROPERTIES WITH APPLICATIONS

NITIS MUKHOPADHYAY AND WILLIAM DUGGAN

Department of Statistics, University of Connecticut, Storrs, CT 06269-3120, U.S.A.

(Received December 3, 1997; revised June 18, 1998)

Abstract. Under a fairly general setup, we first modify the Stein-type two-stage methodology in order to incorporate some *partial information* in the form of a known and positive lower bound for the otherwise *unknown nuisance parameter*, $\theta(> 0)$. This revised methodology is then shown to enjoy various customary *second-order properties and expansions* for functions of the associated stopping variable, under appropriate conditions. Such general machineries are later applied in different types of estimation as well as selection and ranking problems, giving a sense of a very broad spectrum of possibilities. This constitutes natural extensions of these authors' earlier paper (Mukhopadhyay and Duggan (1997*a*, *Sankhyā Ser. A*, **59**, 435–448)) on the fixed-width confidence interval estimation problem exclusively for the mean of a normal distribution having an unknown variance.

Key words and phrases: Consistency, second-order expansions, confidence regions, point estimation, regret, selection and ranking.

1. Introduction

In his classic papers, Stein (1945, 1949) came up with a two-stage procedure for constructing a confidence interval I for estimating the mean μ in a $N(\mu, \sigma^2)$ population when $\sigma^2 \in R^+$ is completely unknown, in such a way that I has the fixed-width $2d$ and $P\{\mu \in I\} \geq 1 - \alpha$ whatever be $\mu \in R$, $\sigma \in R^+$, where $d(> 0)$ and $0 < \alpha < 1$ are preassigned before data collection. The novelty of the two-stage procedure came out loud and clear in showing that the confidence statement was not influenced by the unknown magnitude of the nuisance parameter, σ^2. This was remarkable, particularly because no fixed-sample procedure could solve this problem (Dantzig (1940)) in the first place. The two-stage procedure, unfortunately, has poor characteristics evidenced in oversampling, even asymptotically. In order to circumvent some of these undersirable properties, Mukhopadhyay (1980, 1982) proposed a type of modification in the methodology. The investigations eventually led to a concept, such as *second-order efficiency*, in Ghosh and Mukhopadhyay (1981). Cox (1952) is very pertinent in this context too. On the other hand, purely sequential procedures practically took off from Ray (1957) and Chow and Robbins

(1965) in a natural progression, culminating into Woodroofe's (1977) nonlinear renewal theoretic results. For an overview, the reader is referred to Ghosh *et al.* (1997).

Recently, Mukhopadhyay and Duggan (1997*a*) have considered a modification of Stein's two-stage fixed-width confidence interval estimation procedure for μ in a $N(\mu, \sigma^2)$ population, when the experimenter has some prior knowledge to justify that $\sigma > \sigma_L$ and $\sigma_L(> 0)$ is known. The corresponding two-stage procedure still enjoyed the exact consistency property and the methodology also had attractive *second-order* characteristics associated with it.

In this paper, we focus on two goals. First, in Section 2, we propose a general two-stage procedure, synthesizing and extending the important ideas from Mukhopadhyay and Duggan (1997*a*). The emphasis here lies first in obtaining *second-order* results under a general setup. Sections 3 and 4 provide a variety of explicit applications of our proposed general theory. Section 3 includes applications in estimation of (i) the mean vector of a multivariate normal distribution, (ii) the location in a negative exponential distribution, (iii) the regression parameters in a linear model. Section 4 includes applications in multiple decision theory, for example, in the contexts of selecting the best normal or best negative exponential population. In all these diverse examples, we aim at showing how easily the general theory from Section 2 can be put to work so that we can conclude associated *second-order properties* in specific cases. This, in turn, points toward significant breadth achieved by means of the proposed generalization.

Mukhopadhyay and Duggan (1997*a*) reported encouraging findings obtained from extended sets of simulations carried out in the context of fixed-width confidence interval construction for the mean of a normal population. We have available encouraging moderate sample-size performances of the proposed two-stage sampling design in the case of few examples discussed in this paper. In order to keep our presentation short and crisp, we refrain from including long sets of tables summarizing the findings obtained via computer simulations.

2. General formulation and analyses

In many problems in the areas of estimation and multiple decision theory, the expression of the so called "optimal" fixed sample size turns out to be $n_0^* = q\theta^\tau h^{*-1}$, where q, τ and h^* are known positive numbers, but θ is the unknown and positive nuisance parameter and yet *we are assured that* $\theta > \theta_L(> 0)$ with θ_L known. The explicit roles of q, τ, h^* and θ would be clear from specific applications discussed in Sections 3 and 4. In this section, the asymptotic analyses would be carried out when $h^* \to 0$. Throughout, $[u]^*$ and $I(\cdot)$ would respectively stand for the largest integer $< u$ and the indicator function of $(\cdot)$.

Let $m(\geq m_0)$ be the initial sample size where

$$m \equiv m(h^*) = \max\left\{m_0, \left[\frac{q\theta_L^\tau}{h^*}\right]^* + 1\right\}, \tag{2.1}$$

$m_0(\geq 1)$ being a fixed integer. Based on the pilot sample of size m, suppose that one considers a statistic $U(m)$ so that $P\{U(m) > 0\} = 1$, $E\{U(m)\} = \theta$. In fact, *we require that the estimator $U(m)$ of θ satisfy the following crucial property:*

(2.2) $p_m U(m)/\theta$ is distributed as $\chi^2_{p_m}$ where p_m is a positive integer of the form $c_1 m + c_2$ with positive integer c_1 and integer c_2.

Observe that when (2.2) is satisfied, $U(m) \xrightarrow{P} \theta$ as $h^* \to 0$. Let q^* be positive where

$$q^* \equiv q^*_m = q + c_3 m^{-1} + O(m^{-2}), \tag{2.3}$$

with some real number c_3. Define a positive integer valued random variable as follows:

$$N \equiv N(h^*) = \max\left\{m, \left[\frac{q^*_m U^\tau(m)}{h^*}\right]^* + 1\right\}. \tag{2.4}$$

In various applications, what one does is to start with m random samples in the first stage and obtain $U(m)$, which leads to N that estimates n^*_0. If $N = m$, one does not take any more observations at the second stage. But, if $N > m$, then one samples the difference $(N - m)$ at the second stage. In either situation, one proceeds with the appropriate inference procedures, given the nature of a particular application, that depend on the totality of all N observations obtained from the two combined stages of sampling. The two-stage sampling scheme (2.4) is a generalized version of Mukhopadhyay and Duggan's (1997*a*) strategy. At this point, we set out to derive *second-order* characteristics of N given by (2.4) as $h^* \to 0$, when m is defined by (2.1). Motivations behind these results would be clear from the different applications addressed in Sections 3 and 4.

LEMMA 2.1. *For m and N respectively defined in* (2.1) *and* (2.4), *we have as $h^* \to 0$:*

$$P(N = m) = O(\eta^{(1/2)p_m})$$

where $\eta = (\theta_L/\theta)\exp\{1 - (\theta_L/\theta)\}$ which is a positive proper fraction and p_m is defined in (2.2).

From (2.4), we observe the basic inequality:

$$\frac{q^*_m U^\tau(m)}{h^*} \le N \le mI(N = m) + \frac{q^*_m U^\tau(m)}{h^*} + 1. \tag{2.5}$$

Since $U(m) \xrightarrow{P} \theta$ as $h^* \to 0$ in view of (2.2), $I(N = m) \xrightarrow{P} 0$ as $h^* \to 0$ in view of Lemma 2.1, from (2.5) we can immediately conclude that $N/n^*_0 \xrightarrow{P} 1$ as $h^* \to 0$. Utilizing (2.2), we write $E\{U^s(m)\} = (2\theta/p_m)^s \Gamma(\frac{1}{2}p_m + s)\{\Gamma(\frac{1}{2}p_m)\}^{-1}$, and hence,

$$E\{U^s(m)\} = \theta^s\{1 + s(s-1)(c_1 m)^{-1} + O(m^{-2})\}. \tag{2.6}$$

Equation (2.6) holds whether s is positive or negative, but for s negative, p_m has to exceed $-2s$. Taking expectations throughout (2.5), in view of (2.6) with $s = \tau$ and Lemma 2.1, and then dividing all sides by n^*_0, we can claim that $E(N/n^*_0) \to 1$ as $h^* \to 0$.

2.1 *Second-order analyses*

THEOREM 2.1. *For m and N respectively defined in* (2.1) *and* (2.4), *we have as $h^* \to 0$:*

$$\psi + o(h^{*(1/2)}) \leq E(N) - n_0^* \leq \psi + 1 + o(h^{*(1/2)})$$

where $\psi = q^{-1}\{\tau(\tau-1)qc_1^{-1} + c_3\}(\theta\theta_L^{-1})^\tau$ *and* $n_0^* = q\theta^\tau/h^*$.

This result proves the *second-order efficiency property* of the two-stage procedure in the sense of Ghosh and Mukhopadhyay (1981). The following result is needed in the proof of Theorem 2.1 and it is also of independent interest.

LEMMA 2.2. *For m and N respectively defined in* (2.1) *and* (2.4), *we have:*
(i) $n_0^{*-1/2}(N - n_0^*) \xrightarrow{L} N(0, \sigma_0^2)$ *as* $h^* \to 0$ *where* $\sigma_0^2 = 2c_1^{-1}\tau^2(\theta\theta_L^{-1})^\tau$;
(ii) $n_0^{*-1}(N - n_0^*)^2$ *is uniformly integrable for* $0 < h^* < h_0^*$ *with sufficiently small* h_0^*.

Suppose that $g : R^+ \to R^+$ is a twice differentiable function such that

(2.7) (i) $g''(x)$ is continuous at $x = 1$,

(2.8) (ii) $|g''(x)| \leq \sum_{i=1}^{r} a_i x^{-b_i}$ for all $x \in R^+$,
where a_i's and b_i's are non-negative.

Next, we set out to provide the second-order bounds for $E[g(N/n_0^*)]$ under certain conditions on $g(\cdot)$. As a special case, we would then be able to obtain second-order bounds for all negative moments of N/n_0^*.

THEOREM 2.2. *For m and N respectively defined in* (2.1) *and* (2.4), *with $g(\cdot)$ satisfying both* (2.7) *and* (2.8), *we have as $h^* \to 0$:*
(i) $g(1) + n_0^{*-1}\{\psi g'(1) + \frac{1}{2}\sigma_0^2 g''(1)\} + o(n_0^{*-1}) \leq E\{g(N/n_0^*)\} \leq g(1) + n_0^{*-1}\{(\psi+1)g'(1) + \frac{1}{2}\sigma_0^2 g''(1)\} + o(n_0^{*-1})$ *if* $g'(1) > 0$;
(ii) $g(1) + n_0^{*-1}\{(\psi+1)g'(1) + \frac{1}{2}\sigma_0^2 g''(1)\} + o(n_0^{*-1}) \leq E\{g(N/n_0^*)\} \leq g(1) + n_0^{*-1}\{\psi g'(1) + \frac{1}{2}\sigma_0^2 g''(1)\} + o(n_0^{*-1})$ *if* $g'(1) < 0$;
with ψ and σ_0 respectively defined in Theorem 2.1 *and Lemma* 2.2, *with* $n_0^* = q\theta^\tau/h^*$.

THEOREM 2.3. *For m and N respectively defined in* (2.1) *and* (2.4), *for any fixed non-zero real number t, we have as $h^* \to 0$:*
(i) $1 + tn_0^{*-1}\{\psi + \frac{1}{2}(t-1)\sigma_0^2\} + o(n_0^{*-1}) \leq E\{(N/n_0^*)^t\} \leq 1 + tn_0^{*-1}\{\psi + 1 + \frac{1}{2}(t-1)\sigma_0^2\} + o(n_0^{*-1})$ *if* $t > 0$;
(ii) $1 + tn_0^{*-1}\{\psi + 1 + \frac{1}{2}(t-1)\sigma_0^2\} + o(n_0^{*-1}) \leq E\{(N/n_0^*)^t\} \leq 1 + tn_0^{*-1}\{\psi + \frac{1}{2}(t-1)\sigma_0^2\} + o(n_0^{*-1})$ *if* $t < 0$;
with ψ and σ_0 respectively defined in Theorem 2.1 *and Lemma* 2.2, *with* $n_0^* = q\theta^\tau/h^*$.

Lemma 2.1 implies in fact that $m^s P(N = m) \to 0$ for all fixed $s > 0$. In the proofs of Lemma 2.2 and Theorems 2.1–2.3, we did not have to use the full potential of this result. We are aware of some situations where assumption (2.2) fails, but the asymptotic second-order results in the case of analogous two-stage procedures essentially hold, in the presence of weaker rates of convergence of $P(N = m)$ to zero. The paper of Mukhopadhyay (1997) handles such a situation. Even so, the present generalization is broad enough to successfully synthesize many different types of problems in the areas of sequential estimation as well as selection and ranking. Sections 3 and 4 would testify to that.

Remark 2.1. In (2.4), if one replaces $U^\tau(m)$ by $U^\tau(m)f(m)$ where $f(m) = 1 + f_0 m^{-1} + o(m^{-1})$, then all the results in Section 2 go through except that c_3 involved within ψ must then be replaced by $(c_3' + qf_0)$ where $q_m^* = q + c_3' m^{-1} + O(m^{-2})$.

3. Applications in estimation

First we include the estimation problem for the location parameter μ in a negative exponential population with unknown scale parameter $\sigma(> 0)$. Then, we address the fixed-size confidence ellipsoid construction of the mean vector $\boldsymbol{\mu}$ in a $N_p(\boldsymbol{\mu}, \sigma^2 H)$ population where the nuisance parameter $\sigma(> 0)$ is unknown, but H is a known $p \times p$ positive definite (p.d.) matrix. We also consider the point estimation problems for $\boldsymbol{\mu}$ with regard to the same population. In the end, we briefly indicate an example from the area of linear regression. For a general review, one should consult Ghosh *et al.* (1997).

3.1 *Negative exponential location*

Let $X_1, X_2, \ldots$ be i.i.d. random variables with the probability density function:

$$f(x; \mu, \sigma) = \sigma^{-1} \exp\{-(x - \mu)/\sigma\} I(x > \mu), \tag{3.1}$$

where $-\infty < \mu < \infty$, $0 < \sigma < \infty$ are two unknown parameters. This distribution has been used widely in reliability as well as survival analyses. Sequential and multistage estimation problems for μ and σ were reviewed in Mukhopadhyay (1988, 1995). Having recorded $X_1, \ldots, X_n$ we estimate μ and σ respectively by $T_n = \min\{X_1, \ldots, X_n\}$ and $S_n = (n-1)^{-1} \Sigma_{i=1}^n (X_i - T_n)$, with $n \geq 2$. Now suppose that given two preassigned numbers $d(> 0)$ and $0 < \alpha < 1$, we wish to construct a confidence interval I for μ such that the length of I is d and $P\{\mu \in I\} \geq 1 - \alpha$. Here, the scale parameter is treated as a nuisance parameter. A two-stage procedure was originally proposed by Ghurye (1958), and it does not have the (first-order) efficiency property in the sense of Chow and Robbins (1965) and Ghosh and Mukhopadhyay (1981).

Let us consider $I_n = [T_n - d, T_n]$ as the confidence interval for μ. Now, $P\{\mu \in I_n\} \geq 1 - \alpha$ provided that n is the smallest integer $\geq a\sigma/d = C$, say, where $a = \ln(1/\alpha)$. Now, C plays the role of n_0^* with $q = a$, $\theta = \sigma$, $\tau = 1$ and $h^* = d$. But, *let us suppose that* $\sigma > \sigma_L$ *where* $\sigma_L(> 0)$ *is available from prior knowledge and the nature of the practical applications on hand.* With $\theta_L = \sigma_L$

and $m_0 \geq 2$, one then defines m as in (2.1) and N as in (2.4) with $U(m) = S_m$, and implements the two-stage sampling design, with q_m^* being the upper $100\alpha\%$ point of the F-distribution with degrees of freedom 2 and $2(m-1)$. Based on all the observations $X_1, \ldots, X_N$, we then propose the fixed-width confidence interval $I_N = [T_N - d, T_N]$ for μ.

Since $I(N = n)$ and T_n are independent for all $n \geq m$, we have

$$P\{\mu \in I_N\} = E[1 - \exp(-Nd/\sigma)] = E[g(N/C)] \tag{3.2}$$

where $g(x) = 1-\exp(-ax)$, $x > 0$. The condition (2.2) is satisfied with $p_m = 2(m-1)$, that is $c_1 = 2$, $c_2 = -2$. It is also easy to see that $\alpha = \{1+q_m^*(m-1)^{-1}\}^{-(m-1)}$ and hence $q_m^* = a + \frac{1}{2}a^2m^{-1} + O(m^{-2})$ so that (2.3) holds with $c_3 = \frac{1}{2}a^2$.

From Ghurye (1958), and Mukhopadhyay (1988), it follows that for all fixed μ, σ, d and α,

$$P\{\mu \in I_N\} \geq 1 - \alpha \quad \text{[Consistency Property]}. \tag{3.3}$$

Using Theorem 2.1, we immediately claim that

$$\frac{1}{2}a\sigma\sigma_L^{-1} + o(d^{1/2}) \leq E(N) - C \leq \frac{1}{2}a\sigma\sigma_L^{-1} + 1 + o(d^{1/2}), \tag{3.4}$$

as $d \to 0$, and this refers to the *second-order efficiency property* in the sense of Ghosh and Mukhopadhyay (1981).

Next, we should look for a second-order expansion, more specifically, bounds for the coverage probability. Observe that $|g''(x)| \leq a^2$ for all $x > 0$, $g'(1) = ae^{-a}$, $g''(1) = -a^2e^{-a}$, and (2.7) and (2.8) hold. We should then right away use part (i) of Theorem 2.2 where $\sigma_0^2 = \sigma\sigma_L^{-1}$, and obtain

$$(1 - \alpha) + o(d) \leq P\{\mu \in I_N\} \leq (1 - \alpha) + \alpha \ln(1/\alpha)C^{-1} + o(d) \tag{3.5}$$

as $d \to 0$.

3.2 *Multivariate normal mean vector*

Consider $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$, a sequence of independent $N_p(\boldsymbol{\mu}, \sigma^2 H)$ random variables where $\boldsymbol{\mu} \in R^p$, $\sigma \in R^+$ are two unknown parameters, but H is a $p \times p$ p.d. matrix. Here, σ^2 is the nuisance parameter. First, we address the fixed-size confidence region problem and then the minimum as well as the bounded risk point estimation problems for the mean vector $\boldsymbol{\mu}$. *In practice, let us suppose, however, that $\sigma > \sigma_L$ where $\sigma_L (> 0)$ is known.*

3.2.1 *Fixed-size confidence region*

Having recorded $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, we estimate $\boldsymbol{\mu}$ and σ^2 respectively by $\bar{\boldsymbol{X}}_n = n^{-1}\Sigma_{i=1}^n \boldsymbol{X}_i$ and $S_n^2 = (np-p)^{-1}\Sigma_{i=1}^n(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)'H^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)$ with $n \geq 2$. Given $d(>0)$ and $0 < \alpha < 1$ we consider the fixed-size ellipsoidal confidence region

$$\mathcal{R}_n = \{\boldsymbol{\omega} \in R^p : (\bar{\boldsymbol{X}}_n - \boldsymbol{\omega})'H^{-1}(\bar{\boldsymbol{X}}_n - \boldsymbol{\omega}) \leq d^2\} \tag{3.6}$$

for $\boldsymbol{\mu}$, and we require that $P\{\boldsymbol{\mu} \in \mathcal{R}_n\} \geq 1 - \alpha$ which holds if n is the smallest integer $\geq a\sigma^2/d^2 = C$, say. Here, $F(a) = 1 - \alpha$ with $F(x) = P(\chi_p^2 \leq x)$, $x > 0$. Here again, C plays the role of n_0^* where $q = a$, $h^* = d^2$, $\theta = \sigma^2$ and $\tau = 1$. Mukhopadhyay and Al-Mousawi (1986) proposed a two-stage procedure which had the consistency property but it was not even first-order efficient. Mukhopadhyay and Al-Mousawi (1986) also had developed other multistage procedures. Nagao (1996) came up with the second-order properties associated with his sequential procedure when the dispersion matrix has some special structure, including the one considered in Mukhopadhyay and Al-Mousawi (1986).

With $\theta_L = \sigma_L^2$ and $m_0 \geq 2$, we then define m as in (2.1) and N as in (2.4) with $U(m) = S_m^2$, and implement the two-stage sampling design with $q_m^* = pb_m$ where b_m is the upper $100\alpha\%$ point of the F-distribution with degrees of freedom p and $p(m-1)$. The condition (2.2) is satisfied with $p_m = pm - p$, that is with $c_1 = p = -c_2$. Utilizing the results from Scheffe and Tukey (1944), we can write $b_m = ap^{-1}\{1 - \frac{1}{2}(p-2-a)(pm)^{-1} + O(m^{-2})\}$, and thus (2.3) holds with $c_3 = -\frac{1}{2}a(p-2-a)p^{-1}$.

Since $I(N = n)$ and $\bar{\boldsymbol{X}}_n$ are independent for all $n \geq m$, we have

$$\begin{aligned} P\{\boldsymbol{\mu} \in \mathcal{R}_N\} &= E[F(Nd^2/\sigma^2)] \\ &= E[g(N/C)] \end{aligned} \tag{3.7}$$

with $g(x) = F(ax)$, $x > 0$ while the confidence region $\mathcal{R}_N$ corresponds to (3.6) based on $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$.

From Theorem 2 in Mukhopadhyay and Al-Mousawi (1986) it follows that

$$P\{\boldsymbol{\mu} \in \mathcal{R}_N\} \geq 1 - \alpha \quad \text{[Consistency Property]}, \tag{3.8}$$

for all fixed $\boldsymbol{\mu}$, σ, d and α. Using Theorem 2.1, we also immediately claim that

$$\begin{aligned} -\frac{1}{2}(p-2-a)p^{-1}\sigma^2\sigma_L^{-2} + o(d) &\leq E(N) - C \\ &\leq -\frac{1}{2}(p-2-a)p^{-1}\sigma^2\sigma_L^{-2} + 1 + o(d), \end{aligned} \tag{3.9}$$

as $d \to 0$, while this refers to the *second-order efficiency property*. This matches with the answer provided by Mukhopadhyay and Duggan (1997a) when $p = 1$.

Let us write $h(x;p) = e^{-(1/2)x}x^{(1/2)p-1}$, $x > 0$. Then, $h(x;p)$ has the maximum at $x = x^*(p) = p - 2$ for every fixed $p > 2$. Also, one can verify that $F''(x) = -\{2^{(1/2)p+1}\Gamma(\frac{1}{2}p)\}^{-1}h(x;p) + (\frac{1}{2}p - 1)\{2^{(1/2)p}\Gamma(\frac{1}{2}p)\}^{-1}h(x;p-2)$, and $F'(a) > 0$. In the case $p \geq 4$, $|F''(x)|$ is bounded and hence so is $|g''(x)|$, $x > 0$.

That is, (2.7) and (2.8) are in order. The case $p = 2$ is similar. When $p = 1$ or 3, the upper bound of $|g''(x)|$ would not exactly satisfy the condition (2.8), because in that upper bound, one would see both positive and negative powers of x. An appropriate combination of parts of the proofs of Theorems 2.2 and 2.3 would easily indicate that the conclusions of Theorem 2.2 would still hold in the present situation. Thus, since $\sigma_0^2 = 2p^{-1}\sigma^2\sigma_L^{-2}$, we can immediately conclude that

$$(3.10) \qquad (1-\alpha) + o(d^2) \le P\{\boldsymbol{\mu} \in \mathcal{R}_N\} \le (1-\alpha) + aF'(a)C^{-1} + o(d^2),$$

as $d \to 0$. In order to derive (3.10) explicitly, the relationships such as $F''(a) = F'(a)\{-\frac{1}{2} + \frac{1}{2}(p-2)a^{-1}\}$, $g'(1) = aF'(a)$, $g''(1) = a^2F''(a)$ would be helpful.

3.2.2 *Minimum risk estimation*

Let $\boldsymbol{X}$'s be i.i.d. $N_p(\boldsymbol{\mu}, \sigma^2 H)$ as before. Suppose that the loss function in estimating $\boldsymbol{\mu}$ by $\bar{\boldsymbol{X}}_n$ is taken to be

$$(3.11) \qquad L_n = A\{(\bar{\boldsymbol{X}}_n - \boldsymbol{\mu})'H^{-1}(\bar{\boldsymbol{X}}_n - \boldsymbol{\mu})\}^{(1/2)r} + cn^t$$

where A, c, r and t are known positive numbers. The type of loss function given by (3.11) was adopted by Wang (1980). The situation when $r = 2$ and $t = 1$ corresponds to the customary scenario of squared error loss plus linear cost of sampling. Now, the risk associated with (3.11) is given by $R_n = E(L_n) = B\sigma^r n^{-(1/2)r} + cn^t$ with $B = 2^{(1/2)r}A\Gamma(\frac{1}{2}(p+r))/\Gamma(\frac{1}{2}p)$, whereas this risk is minimized (approximately) if $n = n^* = \{K^*c^{-1}\}^{2/(2t+r)}\sigma^{2r/(2t+r)}$ with $K^* = \frac{1}{2}rBt^{-1}$. We tacitly assume that n^* is an integer. The corresponding minimum risk is given by $R_{n^*} = B\sigma^r n^{*-(1/2)r} + cn^{*t} = cn^{*t}\{1 + BK^{*-1}\} = c(1 + 2tr^{-1})n^{*t}$, and our goal is to achieve this minimum risk approximately.

Note that n^* plays the role of n_0^* where $q = K^{*2/(2t+r)}$, $h^* = c^{2/(2t+r)}$, $\theta = \sigma^2$ and $\tau = r/(2t+r)$. With $\theta_L = \sigma_L^2$ and $m_0 \ge 2$, we then define m as in (2.1) and N as in (2.4) where $U(m) = S_m^2$, the same as in Section 3.2.1, and $q_m^* = q$. The condition (2.2) is satisfied with $c_1 = -c_2 = p$, and (2.3) holds with $c_3 = 0$. After implementing the two-stage procedure, we propose to estimate $\boldsymbol{\mu}$ by $\bar{\boldsymbol{X}}_N$. Using Theorem 2.1, we immediately obtain the following result:

$$(3.12) \qquad \begin{aligned} &-2rt\{p(2t+r)^2\}^{-1}(\sigma^2/\sigma_L^2)^{r/(2t+r)} + o(c^{1/(2t+r)}) \\ &\le E(N) - n^* \\ &\le -2rt\{p(2t+r)^2\}^{-1}(\sigma^2/\sigma_L^2)^{r/(2t+r)} + 1 + o(c^{1/(2t+r)}), \end{aligned}$$

as $c \to 0$, and this refers to the *second-order efficiency property.*

Since $I(N = n)$ and $\bar{\boldsymbol{X}}_n$ are independent for all $n \ge m$, we have $E(L_N) = B\sigma^r E(N^{-(1/2)r}) + cE(N^t)$, and hence the

$$(3.13) \qquad \begin{aligned} \text{Regret} &= E(L_N) - R_{n^*} \\ &= c\{2tr^{-1}[E((N/n^*)^{-(1/2)r}) - 1] + [E((N/n^*)^t) - 1]\}n^{*t}. \end{aligned}$$

Then, utilizing Theorem 2.3, we can immediately obtain the following result: As $c \to 0$,

$$(3.14) \qquad ctn^{*t-1}(\rho - 1) + o(cn^{*t-1}) \le \text{Regret} \le ctn^{*t-1}(\rho + 1) + o(cn^{*t-1}),$$

where $\rho = \frac{1}{2}r^2\{p(2t+r)\}^{-1}(\sigma^2/\sigma_L^2)^{r/(2t+r)}$.

In the case when $t = 1$ and $r = 2$, that is under the *squared error loss plus linear cost*, for all p, the regret given by (3.13) reduces to $cE[(N-n^*)^2/N]$, and hence one can fall back on Lemmas 2.1 and 2.2 directly, without appealing to Theorem 2.3. The final result then would become: Regret $= 2c_1^{-1}\tau^2(\sigma^2/\sigma_L^2)^\tau c + o(c)$, which reduces to $(2p)^{-1}(\sigma/\sigma_L)c + o(c)$. This result was derived in Theorem 4.1 of Mukhopadhyay and Duggan (1997*a*) when $p = 1$. Observe that "ρ" in (3.14) reduces to $(2p)^{-1}(\sigma/\sigma_L)c$ as well when $r = 2$ and $t = 1$, but (3.14) gives the second-order bounds for the regret, rather than the second-order expansion of the regret itself.

Remark 3.1. The corresponding bounded risk point estimation problem for the mean vector was briefly discussed in the technical report of Mukhopadhyay and Duggan (1997*b*). We omit it from here for brevity.

3.3 *Linear regression problem*

Consider the linear regression model with normally distributed errors. We write

$$Y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, 2, \ldots$$

where ε_i's are i.i.d. $N(0,\sigma^2)$, $\boldsymbol{\beta}$ is an unknown $p \times 1$ vector of parameters, and $\boldsymbol{x}_i$'s are known vectors. Let us denote $\boldsymbol{Y}_n' = (Y_1, \ldots, Y_n)$ and $X_n' = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ and assume that the model is of full rank, that is the rank of the $p \times p$ matrix $X_n'X_n$ is $p(< n)$. Also, we assume that the nuisance parameter $\sigma(> 0)$ is unknown.

Having recorded $(\boldsymbol{x}_i, Y_i)$, $i = 1, \ldots, n$, we estimate $\boldsymbol{\beta}$ by the least squares estimator $\hat{\boldsymbol{\beta}}_n = (X_n'X_n)^{-1}X_n'\boldsymbol{Y}_n$ and use the loss function

$$L_n = A(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})'(n^{-1}X_n'X_n)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \tag{3.15}$$

with $A(> 0)$ known. Our goal is to make the risk $\leq W$ where $W(> 0)$ is preassigned. Hence, the sample size n has to be the smallest integer $\geq Ap\sigma^2/W = n^*$, say, which corresponds to n_0^* with $q = Ap$, $h^* = W$, $\tau = 1$ and $\theta = \sigma^2$. *Let us assume that* $\sigma > \sigma_L$ *where* $\sigma_L(> 0)$ *is known.* Then, we implement the two-stage procedure (2.4) where $U(m) = S_m^2$, the mean square error, namely $(m-p)^{-1}\{(\boldsymbol{Y}_m - X_m\hat{\boldsymbol{\beta}}_m)'(\boldsymbol{Y}_m - X_m\hat{\boldsymbol{\beta}}_m)\}$, m given by (2.1) with $\theta_L = \sigma_L^2$ and $m_0 > p+2$. With $q_m = \frac{1}{2}(m-p)[\Gamma\{\frac{1}{2}(m-p-2)\}/\Gamma\{\frac{1}{2}(m-p)\}]$, let $q_m^* = Apq_m$. One can verify that the risk, $E(L_N) = Ap\sigma^2E(N^{-1})$ is at most W, for all fixed $\boldsymbol{\beta}$, σ^2, A and W. This refers to the *risk efficiency property*. Now, (2.1) holds with $p_m = m - p$, that is with $c_1 = 1$, $c_2 = -p$. Again, from 6.1.47 of Abramowitz and Stegun ((1972), p. 257), it follows that $q_m = 1 + 2m^{-1} + O(m^{-2})$, that is (2.3) holds with $c_3 = 2Ap$. Hence, Theorem 2.3 (ii) would provide second-order bounds (as $W \to 0$) for the risk, $E(L_N)$. Further details are omitted.

In this setup, minimum risk point estimation problem for $\boldsymbol{\beta}$ or the fixed-size confidence region problem for $\boldsymbol{\beta}$ could also be easily introduced under similar sort of two-stage sampling schemes when $\sigma > \sigma_L$, with $\sigma_L(> 0)$ known. In order to review such procedures when $\sigma(> 0)$ is completely unknown, one should refer to Ghosh *et al.* (1997), Mukhopadhyay and Abid (1986), and Mukhopadhyay (1991). Finster's (1983, 1985) papers are also very relevant. We omit the details for brevity.

4. Applications in multiple decision theory

The basic theory developed in Section 2 is now applied for two interesting selection and ranking problems. Again the emphasis lies in achieving *asymptotic second-order characteristics* for the newly proposed two-stage methodologies in such problems.

4.1 *Selecting the best normal population*

Consider independent populations $\pi_1, \ldots, \pi_k$ with $k \geq 2$, and suppose that $X_{i1}, \ldots, X_{in}, \ldots$ are i.i.d. $N(\mu_i, \sigma^2)$ random variables from π_i, with $\mu_i \in R$ and $\sigma \in R^+$, $i = 1, \ldots, k$. Let us write $\bar{X}_{in} = n^{-1}\Sigma_{j=1}^n X_{ij}$, $U_{in}^2 = (n-1)^{-1}\Sigma_{j=1}^n (X_{ij} - \bar{X}_{in})^2$, and $U_n = k^{-1}\Sigma_{i=1}^k U_{in}^2$ for $n \geq 2$, $i = 1, \ldots, k$. We assume that all the parameters are unknown whereas σ^2 is considered the nuisance parameter. Let us denote $\boldsymbol{\mu}' = (\mu_1, \ldots, \mu_k)$ and write $\mu_{[1]} \leq \cdots \leq \mu_{[k-1]} \leq \mu_{[k]}$ for the ordered μ values. Pursuing Bechhofer's (1954) *indifference zone formulation*, let there be two preassigned numbers $\delta^*(> 0)$ and $P^* \in (k^{-1}, 1)$, and our goal is to select the population associated with $\mu_{[k]}$, referred to as the *best population*, so that $P(CS) \geq P^*$ whenever $\boldsymbol{\mu} \in \Omega(\delta^*)$, with $\Omega(\delta^*) = \{\boldsymbol{\mu} : \mu_{[k]} - \mu_{[k-1]} \geq \delta^*\}$, the *preference zone*. The parameter subspace $\Omega^c(\delta^*)$ is referred to as the *indifference zone*. Here and elsewhere, "CS" will stand for "Correct Selection". Define $C = h^2\sigma^2/\delta^{*2}$ where "h" satisfies the integral equation: $\int_{-\infty}^{\infty} \Phi^{k-1}(y+h)\phi(y)dy = P^*$, $\phi(\cdot)$ being the standard normal density and $\Phi(x) = \int_{-\infty}^{x} \phi(y)dy$, $x \in R$. If σ^2 were known, then C could be interpreted as the optimal fixed sample size required from each π in conjunction with the selection of the population giving rise to the largest sample mean. We tacitly assume that C is an integer. When $\sigma^2(> 0)$ is completely unknown, a two-stage procedure was developed by Bechhofer *et al.* (1954). For a review of other multistage sampling techniques in this problem, refer to Chapter 3 in Mukhopadhyay and Solanky (1994).

Let us, however, assume that $\sigma > \sigma_L$ where $\sigma_L(> 0)$ is known in advance. Now, C plays the role of n_0^* with $q = h^2$, $\theta = \sigma^2$, $\tau = 1$ and $h^* = \delta^{*2}$. One then defines m as in (2.1), with $m_0 \geq 2$, $\theta_L = \sigma_L^2$, and considers N as in (2.4) with $U(m) = U_m$ and $q_m^* = q$. We then implement the two-stage methodology and select the population associated with $\max_{1 \leq i \leq k} \bar{X}_{iN}$ based on the observations $\{X_{i1}, \ldots, X_{iN}, i = 1, \ldots, k\}$. Since $I(N = n)$ and $(\bar{X}_{1n}, \ldots, \bar{X}_{kn})$ are independent for all $n \geq m$, we have (from Theorem 3.2.1 in Mukhopadhyay and Solanky (1994)):

$$
(4.1) \qquad \inf_{\mu \in \Omega(\delta^*)} P(CS) = E\left[\int_{-\infty}^{\infty} \{\Phi(y + N^{1/2}\delta^*\sigma^{-1})\}^{k-1}\phi(y)dy\right] = E[g(N/C)]
$$

where $g(x) = f(x^{1/2})$ and $f(x) = \int_{-\infty}^{\infty} \Phi^{k-1}(y + hx)\phi(y)dy$, $x > 0$. The condition (2.2) is satisfied with $p_m = k(m-1)$, that is $c_1 = -c_2 = k$. Obviously (2.3) holds with $c_3 = 0$.

Using Theorem 2.1, we immediately claim that as $\delta^* \to 0$,

$$
(4.2) \qquad o(\delta^*) \leq E(N) - C \leq 1 + o(\delta^*),
$$

since ψ, defined in Theorem 2.1, turns out to be zero, and this refers to the *second-order efficiency property*. Next, we look at the second-order bounds for the probability of correct selection under the least favorable configuration, that is when $\mu_{[1]} = \cdots = \mu_{[k-1]} = \mu_{[k]} - \delta^*$, where $P(CS)$ attains its infimum for $\boldsymbol{\mu} \in \Omega(\delta^*)$. The conditions (2.7) and (2.8) hold with $r = 3$, $b_1 = \frac{1}{2}$, $b_2 = 1$, and $b_3 = \frac{3}{2}$. Since $g'(1)$ is positive, Theorem 2.2 (i) leads us to claim that

$$\text{(4.3)} \qquad \begin{aligned} &P^* + k^{-1}(\sigma^2/\sigma_L^2)g''(1)C^{-1} + o(\delta^{*2}) \\ &\le \inf_{\boldsymbol{\mu}\in\Omega(\delta^*)} P(CS) \\ &\le P^* + \{k^{-1}(\sigma^2/\sigma_L^2)g''(1) + g'(1)\}C^{-1} + o(\delta^{*2}), \end{aligned}$$

as $\delta^* \to 0$. Observe that $g(1) = P^*$, and σ_0^2, defined in Lemma 2.2 simplifies to $2k^{-1}\sigma^2\sigma_L^{-2}$.

Remark 4.1. One should note that in order to mimic the original two-stage procedure of Bechhofer *et al.* (1954), we could instead use $q_m^* \equiv \tau_1^2(m, k, P^*)$ where $P\{T_i \le 2^{-1/2}\tau_1, i = 1, \ldots, k-1\} = P^*$, with $(T_1, \ldots, T_{k-1})$ distributed as $(k-1)$ dimensional multivariate t with equicorrelation $= \frac{1}{2}$ and the degrees of freedom $= k(m-1)$. In that case the two-stage procedure (2.4) would have the consistency property, namely, that $P(CS) \ge P^*$ for all fixed $\boldsymbol{\mu} \in \Omega(\delta^*)$, σ^2, P^* and δ^*. If such q_m^* satisfies (2.3), then we can easily obtain results analogous to those in (4.2) and (4.3), and these obviously hold when $k = 2$ and $q_m^* = \tau_1^2$.

4.2 *Selecting the best negative exponential population*

Consider independent populations $\pi_1, \ldots, \pi_k$ with $k \ge 2$, and suppose that $X_{i1}, \ldots, X_{in}, \ldots$ are i.i.d. random variables obtained from π_i having the probability density function $f(x; \mu_i, \sigma)$, defined via (3.1), with $\mu_i \in R$, $\sigma \in R^+$, $i = 1, \ldots, k$. Let us write $T_{in} = \min\{X_{i1}, \ldots, X_{in}\}$, $U_{in} = (n-1)^{-1}\Sigma_{j=1}^n (X_{ij} - T_{in})$, and $U_n = k^{-1}\Sigma_{i=1}^k U_{in}$ for $n \ge 2$, $i = 1, \ldots, k$. We assume that all the parameters are unknown whereas σ is considered the nuisance parameter. Let us denote as before $\boldsymbol{\mu}$, $\mu_{[\cdot]}$, and pursue the *indifference zone formulation* again, given two preassigned numbers $\delta^*(> 0)$ and $P^* \in (k^{-1}, 1)$. We define the *preference zone* $\Omega(\delta^*)$ as before and the problem is to select the population associated with $\mu_{[k]}$, referred to as the *best population*, in such a way that $P(CS) \ge P^*$ whenever $\boldsymbol{\mu} \in \Omega(\delta^*)$. Let $C = a\sigma/\delta^*$ where "a" is obtained by solving the equation, $\int_0^\infty \{1 - \exp(-z - a)\}^{k-1} \exp(-z)dz = P^*$. If σ were known, then C could be interpreted as the optimal fixed sample size required from each π in conjunction with the selection of the population associated with the largest sample minimum order statistics among the corresponding T_{iC}'s. We tacitly assume that C is an integer. When $\sigma^2(> 0)$ is completely unknown, Desu *et al.* (1977) developed a two-stage procedure for this selection problem. For a review of other multistage sampling techniques in this problem, refer to Chapter 4 in Mukhopadhyay and Solanky (1994), and Panchapakesan (1995).

Let us, however, assume that $\sigma > \sigma_L$ where $\sigma_L(> 0)$ is known in advance. Now, C plays the role of n_0^* where $q = a$, $\theta = \sigma$, $\tau = 1$ and $h^* = \delta^*$. One then defines m as in (2.1), with $m_0 \geq 2$, $\theta_L = \sigma_L$, and considers N as in (2.4) with $U(m) = U_m$ and $q_m^* = q$. We then implement the two-stage methodology and select the population associated with $\max_{1 \leq i \leq k} T_{iN}$ based on the observations $\{X_{i1}, \ldots, X_{iN}\}$, $i = 1, \ldots, k$. Since $I(N = n)$ and $(T_{1n}, \ldots, T_{kn})$ are independent for all $n \geq m$, we have (from Theorem 4.2.1 in Mukhopadhyay and Solanky (1994)):

$$(4.4) \qquad \inf_{\mu \in \Omega(\delta^*)} P(CS) = E\left[\int_0^\infty \{1 - \exp(-z - N\delta^*\sigma^{-1})\}^{k-1} \exp(-z)dz\right] = E[g(N/C)]$$

where $g(x) = \int_0^\infty \{1 - \exp(-z - ax)\}^{k-1} \exp(-z)dz$, $x > 0$. The expression in (4.4) can be further simplified as follows:

$$(4.5) \qquad \inf_{\mu \in \Omega(\delta^*)} P(CS) = \Sigma_{u=0}^{k-1} b(k,u) E[g_u(N/C)]$$

where $b(k,u) = \binom{k-1}{u}(-1)^u(u+1)^{-1}$ and $g_u(x) = \exp(-uax)$, $x > 0$, $u = 0, 1, \ldots, k-1$. The condition (2.2) is satisfied with $p_m = 2k(m-1)$, that is with $c_1 = -c_2 = 2k$. Obviously (2.3) holds with $c_3 = 0$.

Using Theorem 2.1, we immediately claim that

$$(4.6) \qquad o(\delta^{*(1/2)}) \leq E(N) - C \leq 1 + o(\delta^{*(1/2)})$$

as $\delta^* \to 0$, since ψ, defined in Theorem 2.1, turns out to be zero, and this refers to the *second-order efficiency* property. Next, we look at the second-order bounds for the probability of correct selection under the least favorable configuration, that is when $\mu_{[1]} = \cdots = \mu_{[k-1]} = \mu_{[k]} - \delta^*$, where $P(CS)$ attains its infimum for $\mu \in \Omega(\delta^*)$. Combine (4.4) and (4.5) and observe that $|g''(x)|$ is indeed bounded for all $x > 0$, $g(1) = P^*$ and $g'(1)$ is positive. Note that σ_0^2, defined in Lemma 2.2, reduces to $k^{-1}(\sigma/\sigma_L)$, and hence Theorem 2.2 (i) would immediately provide second-order bounds for the expression given in (4.5). Other details are omitted for brevity.

Acknowledgements

Two referees gave helpful remarks and suggestions. One referee was particularly kind in pointing out several important typographical errors within technical matters. We thank both the referees.

Appendix

A.1 *Proof of Lemma* 2.1

For technical convenience, let us choose a sequence of $h^* \to 0$ such that $q\theta_L^\tau h^{*-1}$ always remains an integer $\geq m_0$. Now, for sufficiently small $h^*(> 0)$,

$$\begin{aligned}(A.1) \qquad P(N = m) &= P\{q_m^* U^\tau(m) < h^* m\} \\ &= P\{\chi_{p_m}^2 < (qq_m^{*-1})^{1/\tau}(\theta_L\theta^{-1})p_m\} \\ &\leq \inf_{t>0} E\{\exp(-t\chi_{p_m}^2)\} \exp\{tp_m(qq_m^{*-1})^{1/\tau}(\theta_L\theta^{-1})\} \\ &= (1 + 2t_o)^{-(1/2)p_m} \exp\{t_0 p_m(qq_m^{*-1})^{1/\tau}(\theta_L\theta^{-1})\}\end{aligned}$$

with $t = t_0 = \frac{1}{2}\{(\theta\theta_L^{-1})(q_m^* q^{-1})^{1/\tau} - 1\}$, which can then be made larger than $\frac{1}{2}\{\theta\theta_L^{-1}(1-\varepsilon) - 1\}$ if $h^* \leq h_\varepsilon^*$, for any positive ε. In other words, we can mathematically guarantee that t_0 is positive if $h^* \leq h_\varepsilon^*$ with, say, $\varepsilon = \frac{1}{2}(1 - \theta_L \theta^{-1})$. The result then follows from (A.1), analogously as in Lemma 2.1 in Mukhopadhyay and Duggan (1997*a*). □

A.2 *Proof of Theorem* 2.1

Throughout the basic inequality (2.5) we take expectations and obtain

$$\text{(A.2)} \qquad q_m^* E\{U^\tau(m)\} h^{*-1} \leq E(N) \leq mP(N=m) + q_m^* E\{U^\tau(m)\} h^{*-1} + 1.$$

In view of Lemma 2.1, we have $mP(N = m) = o(h^{*(1/2)})$, and we also use (2.6) with $s = \tau$ and the expansion of q_m^* from (2.3). The result then follows. □

A.3 *Proof of Lemma* 2.2

For technical convenience, let $T = \max\{m, \frac{q_m^* U^\tau(m)}{h^*}\}$ and note that $T \leq N \leq T + 1$. Thus, it will suffice to verify the lemma with N replaced by T. Observe that $P(T = m)$ has the same order given in Lemma 2.1. We start with the basic inequality

$$\text{(A.3)} \qquad \frac{q_m^* U^\tau(m)}{h^*} \leq T \leq mI(T=m) + \frac{q_m^* U^\tau(m)}{h^*},$$

and rewrite it as

$$\text{(A.4)} \qquad V(m) + k_m \leq n_0^{*-1/2}(T - n_0^*) \leq mI(T=m) + V(m) + k_m,$$

where $V(m) = (q_m^*/h^*) n_0^{*-1/2}\{U^\tau(m) - \theta^\tau\}$ and $k_m = (\theta^\tau/h^*) n_0^{*-1/2}(q_m^* - q)$. In view of Lemma 2.1, it is clear that $mI(T = m) \xrightarrow{P} 0$ as $h^* \to 0$. From (2.3), it follows that $k_m \to 0$ as $h^* \to 0$. Next, we utilize (2.2) to claim that $p_m^{1/2}\{U(m) - \theta\} \xrightarrow{L} N(0, 2\theta^2)$ and hence, $p_m^{1/2}\{U^\tau(m) - \theta^\tau\} \xrightarrow{L} N(0, 2\tau^2\theta^{2\tau})$ as $h^* \to 0$. We can rewrite $V(m)$ as $(\theta^\tau \theta_L^\tau c_1)^{-1/2} p_m^{1/2}\{U^\tau(m) - \theta^\tau\} + o_p(1)$ and thus $V(m) \xrightarrow{L} N(0, \sigma_0^2)$ with $\sigma_0^2 = 2c_1^{-1}\tau^2(\theta\theta_L^{-1})^\tau$. This is part (i).

Next, we use (A.3) to write

$$\text{(A.5)} \qquad \begin{aligned} &\frac{q_m^*}{h^*} E\{U^\tau(m)\} \leq E(T) \leq mP(T=m) + \frac{q_m^*}{h^*} E\{U^\tau(m)\}, \\ &\frac{q_m^{*2}}{h^{*2}} E\{U^{2\tau}(m)\} \leq E(T^2) \leq 3m^2 P(T=m) + \frac{q_m^{*2}}{h^{*2}} E\{U^{2\tau}(m)\}, \end{aligned}$$

and combine these with (2.3), and (2.6) for $s = \tau,\, 2\tau$, in order to claim that

$$\begin{aligned} E\{(T - n_0^*)^2\} \geq\; & h^{*-2}\{q^2 + 2qc_3 m^{-1} + o(m^{-1})\}\theta^{2\tau} \\ & \times \{1 + 2\tau(2\tau - 1)(c_1 m)^{-1} + o(m^{-1})\} \\ & - 2n_0^* h^{*-1}\{q + c_3 m^{-1} + o(m^{-1})\}\theta^\tau \\ & \times \{1 + \tau(\tau - 1)(c_1 m)^{-1} + o(m^{-1})\} + n_0^{*2} + o(1). \end{aligned}$$

In other words,

$$\text{(A.6)}\qquad \begin{aligned} E\{n_0^{*-1}(T-n_0^*)^2\} &\geq n_0^*(q^2m)^{-1}\{2qc_3+2q^2\tau(2\tau-1)c_1^{-1}\} \\ &\quad - 2n_0^*(qm)^{-1}\{c_3+q\tau(2\tau-1)c_1^{-1}\}+o(1) \\ &= 2c_1^{-1}\tau^2(\theta\theta_L^{-1})^\tau+o(1). \end{aligned}$$

In a similar fashion, we exploit (A.5) all over again, and we can then show that $E\{n_0^{*-1}(T-n_0^*)^2\} \leq 2c_1^{-1}\tau^2(\theta\theta_L^{-1})^\tau + o(1)$, and combining this with (A.6), it becomes clear that $E\{n_0^{*-1}(T-n_0^*)^2\} \to 2c_1^{-1}\tau^2(\theta\theta_L^{-1})^\tau$ as $h^* \to 0$. Now, part (ii) follows from part (i). □

A.4 *Proof of Theorem* 2.2

We will provide only an outline. For some random variable ξ between 1 and N/n_0^*, we write

$$\text{(A.7)}\qquad E\{g(N/n_0^*)\} = g(1) + n_0^{*-1}\left\{g'(1)E(N-n_0^*) + \frac{1}{2}E[Qg''(\xi)]\right\}$$

with $Q = n_0^{*-1}(N-n_0^*)^2$. For the first term within { } in (A.7), we simply use Theorem 2.1. For the second term within { }, we split the integral over the sets $[N=m]$ and $[N>m]$ and use Lemmas 2.1 and 2.2 to show that $E[Qg''(\xi)] \to \sigma_0^2 g''(1)$ as $h^* \to 0$. We leave out the rest for brevity. □

A.5 *Proof of Theorem* 2.3

Consider $g(x) = x^t$ for $x > 0$ and hence the result follows from Theorem 2.2 when $t < 2$. In the case $t = 2$, $g''(x) = 2$ for all $x > 0$, and hence Lemmas 2.1 and 2.2 and Theorem 2.1 together lead to the result.

Next, let us focus on the case when $t > 2$. With our specific $g(\cdot)$, equation (A.7) reduces to

$$E\{(N/n_0^*)^t\} = 1 + n_0^{*-1}\left\{tE(N-n_0^*) + \frac{1}{2}t(t-1)E[Q\xi^{t-2}]\right\} + o(n_0^{*-1})$$

where ξ lies between 1 and N/n_0^*. The result will follow once we show that

$$\text{(A.8)}\qquad E\{Q\xi^{t-2}\} = \sigma_0^2 + o(1).$$

Let us choose some $\rho > 4$ and note that $\xi < \rho$ on the set $[N \leq \rho n_0^*]$. Thus, $Q\xi^{t-2}I(N \leq \rho n_0^*) \leq \rho^{t-2}Q$, and we use Lemma 2.2 to claim that $E[Q\xi^{t-2}I(N \leq \rho n_0^*)] = \sigma_0^2 + o(1)$. But, $\xi < N/n_0^*$ on the set $[N > \rho n_0^*]$ and hence

$$\text{(A.9)}\qquad \begin{aligned} E\{Q\xi^{t-2}I(N>\rho n_0^*)\} &\leq E\{Q(N/n_0^*)^{t-2}I(N>\rho n_0^*)\} \\ &= E\left\{N^t\left(1-\frac{n_0^*}{N}\right)^2 n_0^{*1-t}I(N>\rho n_0^*)\right\} \\ &\leq n_0^*E\{(N/n_0^*)^t I(N>\rho n_0^*)\} \\ &\leq n_0^*E^{1/2}\{(N/n_0^*)^{2t}\}P^{1/2}(N>\rho n_0^*). \end{aligned}$$

Now, from (A.3) and dominated ergodic theorem one can see that all positive moments of N/n_0^* are finite, and hence from (A.9), we will claim that $E\{Q\xi^{t-2}I(N > \rho n_0^*)\} = o(1)$, once we verify that

$$\text{(A.10)} \qquad n_0^{*2}P(N > \rho n_0^*) \to 0 \quad \text{as} \quad h^* \to 0.$$

Let us assume for technical convenience that ρn_0^* is an integer. Writing $Y^* = mI(N = m)$ and $Z^* = qU^\tau(m)/h^*$, we have the following:

$$\begin{aligned}\text{(A.11)} \quad P(N \geq \rho n_0^* + 1) &\leq P\{Y^* + Z^* q_m^* q^{-1} \geq \rho n_0^*\} \\ &\leq P\left\{Y^* \geq \frac{1}{2}\rho n_0^*\right\} + P\left\{Z^* q_m^* q^{-1} \geq \frac{1}{2}\rho n_0^*\right\} \\ &\leq 2m(\rho n_0^*)^{-1}P(N = m) + P\left\{U(m) \geq \left(\frac{1}{4}\rho\right)^{1/\tau}\theta\right\} \\ &\leq O(\eta^{(1/2)p_m}) + P\{|U(m) - \theta| \geq \varepsilon\},\end{aligned}$$

in view of Lemma 2.1, with $\varepsilon = \theta\{(\frac{1}{4}\rho)^{1/\tau} - 1\}$ which is positive. Using (2.2) and (A.11), we then get for all $s > 0$,

$$\text{(A.12)} \qquad P\{N \geq \rho n_0^* + 1\} \leq O(\eta^{(1/2)p_m}) + O(m^{-s}),$$

which implies (A.10) when we plug in $s > 2$. That is, $E\{Q\xi^{t-2}I(N > \rho n_0^*)\} = o(1)$. Now, the proof of (A.8) is complete. □

References

Abramowitz, M. and Stegun, I. (1972). *Handbook of Mathematical Functions*, Dover, New York.

Bechhofer, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances, *Ann. Math. Statist.*, **25**, 16–39.

Bechhofer, R. E., Dunnett, C. W. and Sobel, M. (1954). A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance, *Biometrika*, **41**, 170–176.

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.

Desu, M. M., Narula, S. C. and Villarreal, B. (1977). A two-stage procedure for selecting the best of k exponential distributions, *Comm. Statist. Theory Methods*, **6**, 1231–1243.

Finster, M. (1983). A frequentist approach to sequential estimation in the general linear model, *J. Amer. Statist. Assoc.*, **78**, 403–407.

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

Ghosh, M. and Mukhopadhyay, N. (1981). Consistency and asymptotic efficiency of two-stage and sequential estimation procedures, *Sankhyā Ser. A*, **43**, 220–227.

Ghosh, M., Mukhopadhyay, N. and Sen, P. K. (1997). *Sequential Estimation*, Wiley, New York.

Ghurye, S. G. (1958). Note on sufficient statistics and two-stage procedures, *Ann. Math. Statist.*, **29**, 155–166.

Mukhopadhyay, N. (1980). A consistent and asymptotically efficient two-stage procedure to construct fixed-width confidence intervals for the mean, *Metrika*, **27**, 281–284.

Mukhopadhyay, N. (1982). Stein's two-stage procedure and exact consistency, *Skandinavisk Aktuarietidskrift*, 110–122.

Mukhopadhyay, N. (1988). Sequential estimation problems for negative exponential populations, *Comm. Statist. Theory Methods* (Reviews Section), **17**, 2471–2506.

Mukhopadhyay, N. (1991). Parametric sequential point estimation, Chapter 10, *Handbook of Sequential Analysis* (eds. B. K. Ghosh and P. K. Sen), 245–267, Marcel Dekker, New York.

Mukhopadhyay, N. (1995). Two-stage and multi-stage estimation, Chapter 26, *The Exponential Distribution: Theory, Methods and Applications* (eds. N. Balakrishnan and A. P. Basu), 429–452, Gordon and Breach Publishers, Amsterdam.

Mukhopadhyay, N. (1997). Second-order properties of a two-stage fixed-size confidence region for the mean vector of a multivariate normal distribution, Statist. Tech. Report No. 97-14, University of Connecticut, Storrs.

Mukhopadhyay, N. and Abid, A. D. (1986). On fixed-size confidence regions for the regression parameters, *Metron*, **44**, 297–306.

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.

Mukhopadhyay, N. and Duggan, W. T. (1997*a*). Can a two-stage procedure enjoy second-order properties?, *Sankhyā Ser. A*, **59**, 435–448.

Mukhopadhyay, N. and Duggan, W. T. (1997*b*). On a two-stage procedure having second-order properties with applications, Statist. Tech. Report No. 97-32, University of Connecticut, Storrs.

Mukhopadhyay, N. and Solanky, T. K. S. (1994). *Multistage Selection and Ranking Procedures*, Marcel Dekker, New York.

Nagao, H. (1996). On fixed width confidence regions for multivariate normal mean when the covariance matrix has some structure, *Sequential Anal.*, **15**, 37–46 (Correction: *ibid.* (1998). **17**, 125–126).

Panchapakesan, S. (1995). Selection and ranking procedures, Chapter 16, *The Exponential Distribution: Theory, Methods and Applications* (eds. N. Balakrishnan and A. P. Basu), 259–278, Gordon and Breach Publishers, Amsterdam.

Ray, W. D. (1957). Sequential confidence intervals for the mean of a normal population with unknown variance, *J. Roy. Statist. Soc. Ser. B*, **19**, 133–143.

Scheffe, H. and Tukey, J. W. (1944). A formula for sample sizes for population tolerance limits, *Ann. Math. Statist.*, **15**, p. 217.

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

Stein, C. (1949). Some problems in sequential estimation (abstract), *Econometrica*, **17**, 77–78.

Wang, Y. H. (1980). Sequential estimation of the mean of a multinormal population, *J. Amer. Statist. Assoc.*, **75**, 977–983.

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