2

Two-Stage and Multistage Procedures

SHOUTIR KISHORE CHATTERJEE
Calcutta University, Calcutta, India

1. INTRODUCTION

Broadly speaking, any sampling procedure in which the number of observations is not predetermined, is determined by the actual observations encountered in the experiment, and is more or less unrestricted, is called a sequential procedure. It follows that in a sequential procedure the observations would be taken not all at a time but in two or more stages. However, by standard usage, only a sequential procedure in which the number of stages (but naturally, not the number of observations in every stage) is bounded is called a multistage procedure.

The object of sequential procedures as developed by Wald (1947) in the 1940s is "avoidance of oversampling." In these procedures, while taking successive observations, one continually monitors the accumulated evidence at every stage; as soon as this is clear-cut, one stops. In this way one is able to realize a prescribed level of performance while keeping the number of observations minimal on an average. Such continuous monitoring with a view to cashing in on one's good luck in getting decisive evidence early is meaningful irrespective of whether the underlying distribution does or does not involve some nuisance parameters. Thus even for the elementary problem of testing a simple hypothesis $\mu = \mu^{(0)}$ against a simple alternative $\mu = \mu^{(1)}$ about the mean of a normal population $N(\mu, 1)$ to realize a given size and power, the SPRT (sequential probability ratio test) is known to achieve on an average a substantial saving in the number of observations over the fixed sample-size procedure. In such a sequential procedure, to avoid oversampling one proceeds, as it were, inch by inch, and, therefore, the number of observations

taken in a stage is either one or a very small number, while the number of stages is generally unlimited. Naturally, such procedures may not be very convenient if the experimental cost has a major component which increases with the number of stages.

At about the same time Wald developed the SPRT, Stein (1945) proposed a kind of sequential inference procedure which is relevant only when the underlying distribution involves, besides the parameter of interest, some nuisance parameter. The object here can be described as *comprehensive studentization*. In such a procedure a pilot sample is taken to estimate the nuisance parameter. The ultimate number of observations is determined on the basis of this estimate, so that the procedure achieves a stipulated level of performance irrespective of the nuisance parameter. These procedures are generally two or multistage procedures and are naturally convenient when the cost of the experiment depends on the number of stages. To illustrate, consider the problem of inference about μ for a population $N(\mu, \sigma^2)$ with both μ and σ^2 unknown. The usual confidence interval for μ, based on a fixed number of observations, attains the prescribed confidence coefficient independently of μ and σ^2, but its span is a random variable whose distribution depends on σ^2. Stein's procedure seeks to construct a confidence interval which achieves not only a stipulated confidence but has a predetermined span. Similarly for testing a hypothesis $\mu = \mu^{(0)}$, the usual test, based on a fixed number of observations, has a predetermined size, but its power involves σ^2 and hence cannot be made to attain a desired value at a specified distance from $\mu^{(0)}$ by planning the experiment. Stein's procedure is geared to meet such a requirement.

Stein proposed procedures to get tests and confidence estimates for a normal mean and for the parameters of a general linear model representing a replicable set of observations. Afterwards, similar procedures were developed for other setups, and the same ideas were used to tackle other problems of inference (Chapman, 1950, Chatterjee 1959a,b, 1962, Ghurye 1958, Healy 1956, Lehmann 1950). We give an account of all these and indicate some further developments. First however, we discuss why it is fruitless to try to construct the type of interval estimate and test envisaged above on the basis of a fixed number of observations.

2. NONEXISTENCE OF FIXED SAMPLE-SIZE PROCEDURES WHOSE PERFORMANCE IS FREE FROM NUISANCE PARAMETERS

As early as 1940, Dantzig (1940) showed that for testing for the normal mean in the presence of an unknown σ^2, no nontrivial test based on a fixed number of observations can have its power free from σ^2. Later, Lehmann (1950) showed that generally for a location-scale family of distributions with an unknown scale parameter, no confidence interval of fixed span can be constructed for the location parameter on the basis of a fixed number of observations. Both these conclusions essentially follow from the fact that two distributions with different locations can be made to approach each other uniformly in measure arbitrarily closely by making the scale parameter sufficiently large.

Specifically, suppose $(\tilde{y}_1, \ldots, \tilde{y}_n)' = \tilde{\mathbf{y}}$ represents a fixed number of random variables whose joint distribution is given by the density function

$$\sigma^{-n} f\left(\frac{y_1 - \mu}{\sigma}, \ldots, \frac{y_n - \mu}{\sigma}\right) = \sigma^{-n} f(\sigma^{-1}(\mathbf{y}_\mu \mathbf{1})), \qquad \mathbf{y} = (y_1, \ldots, y_n)', \ \mathbf{1} = (1, \ldots, 1)'. \tag{2.1}$$

(Here y_k is the running variable corresponding to the random variable $\tilde{y}_k$. Whenever necessary, we distinguish a random variable from the corresponding running variable by a tilde.) We assume $f(\mathbf{x})$ is everywhere continuous in $x \in \mathcal{R}^n$, $-\infty < \mu < \infty$ and $0 < \sigma < \infty$. Then, for any $\mu_1 \neq \mu_2$, as $\sigma \to \infty$,

$$\sup_{\mathcal{E}} \left| \int_{\mathcal{E}} \sigma^{-n} f(\sigma^{-1}(\mathbf{y} - \mu_1 \mathbf{1})) \prod dy_k - \int_{\mathcal{E}} \sigma^{-n} f[\sigma^{-1}(\mathbf{y} - \mu_2 \mathbf{1})] \prod dy_k \right| \longrightarrow 0, \tag{2.2}$$

the supremum being taken over all measurable sets $\mathcal{E} \subset \mathcal{R}^n$. This can be proved by transforming $\mathbf{y}$ to $\mathbf{x} = \sigma^{-1}(\mathbf{y} - \mu_1 \mathbf{1})$ in both the integrals and noting that, as by the continuity of $f(\cdot)$, $f(\mathbf{x} + \sigma^{-1}(\mu_1 - \mu_2)\mathbf{1}) \to f(\mathbf{x})$ for $\sigma \to \infty$, Scheffe's theorem (see e.g., Rao (1973, p. 104) implies (2.2).

Now suppose both μ and σ in (2.1) are unknown and it is possible to find on the basis of $\mathbf{y}$ a confidence set $\mathcal{S}(y) \subset \mathcal{R}$ for μ such that, fixed $0 < 1 - \alpha < 1$, $0 < d < \infty$,

$$P\{\mu \in \mathcal{S}(\tilde{\mathbf{y}}) \mid \mu, \sigma\} \geq 1 - \alpha \qquad \forall \mu, \sigma, \tag{2.3}$$

$$\text{max. diameter } \mathcal{S}(y) < 2d \qquad \forall \mathbf{y} \in \mathcal{R}^n. \tag{2.4}$$

As $-\infty < \mu < \infty$, for any integer $r \geq 2$ we can find r μ-values, $\mu_1, \ldots, \mu_r$, such that the absolute difference between any two of these is $2d$ or more. Then (2.4) implies that the sets $\mathcal{E}_j = \{\mathbf{y}: \mathbf{y} \in \mathcal{R}^n, \mu_j \in \mathcal{S}(y)\}$, $j = 1, \ldots, r$, are mutually exclusive. Again by (2.3),

$$P\{\tilde{\mathbf{y}} \in \mathcal{E}_j \mid \mu_j, \sigma\} \geq 1 - \alpha, \qquad j = 1, \ldots, r,$$

and by (2.2) it is possible to choose σ to ensure the finite number of inequalities

$$\sup_{\mathcal{E}} |P\{\tilde{\mathbf{y}} \in \mathcal{E} | \mu_j, \sigma\} - P\{\tilde{\mathbf{y}} \in \mathcal{E} | \mu_1, \sigma\}| < \tfrac{1}{2}(1 - \alpha), j = 2, \ldots, r.$$

Hence

$$P\{\tilde{\mathbf{y}} \in \mathcal{E}_j \mid \mu_1, \sigma\} > \tfrac{1}{2}(1 - \alpha), \qquad j = 1, 2, \ldots, r. \tag{2.5}$$

Since $\mathcal{E}_j$, $j = 1, \ldots, r$, are mutually exclusive, (2.5) implies $r(1 - \alpha)/2 < 1$. Since $r \geq 2$ is an arbitrary integer, this is clearly impossible. This negates the existence of a confidence set satisfying (2.3)–(2.4).

Similarly in the case of testing $H^{(0)}$: $\mu = \mu^{(0)}$ against $H^{(1)}$: $\mu = \mu^{(1)}$, suppose it is possible to find a test with critical function $\phi(\mathbf{y})$ such that

$$E_{\mu^{(0)}, \sigma} \phi(\tilde{\mathbf{y}}) = \alpha, \quad E_{\mu^{(1)}, \sigma} \phi(\tilde{\mathbf{y}}) = \beta \qquad \forall \sigma \tag{2.6}$$

$0 < \alpha, \beta < 1$. For any fixed σ, if $\mathcal{E}_\sigma \subset \mathcal{R}^n$ is a most powerful critical region of size α for the simple hypothesis specifying $(\mu^{(0)}, \sigma)$ against the simple alternative specifying $(\mu^{(1)}, \sigma)$ as obtained through the Neyman–Pearson lemma, then clearly (2.6) means

$$\beta \leq P\{\tilde{\mathbf{y}} \in \mathcal{E}_\sigma \mid \mu^{(1)}, \sigma\} \qquad \forall \sigma. \tag{2.7}$$

Since $P\{\tilde{\mathbf{y}} \in \mathcal{E}_\sigma \mid \mu^{(0)}, \sigma\} = \alpha \forall \sigma$, by (2.2), as $\sigma \to \infty$,

$$P\{\tilde{\mathbf{y}} \in \mathcal{E}_\sigma \mid \mu^{(1)}, \sigma\} \longrightarrow \alpha. \tag{2.8}$$

Clearly (2.7) and (2.8) imply $\beta \leq \alpha$. Again interpreting $\phi(\cdot)$ as a size β test for $H^{(1)}$ against $H^{(0)}$ with power α independently of σ, by a similar argument we get $\alpha \leq \beta$. Thus (2.6) implies $\alpha = \beta$, so the test $\phi(\cdot)$ is a trivial one.

The above reasoning can be easily extended to more general situations where the role of the scalar location parameter would be taken over by a parameter vector and the single σ would be replaced by multiple nuisance parameters. The main requirement is that it should be possible to obliterate the effect of differences in location on the probabilities of sets in a uniform manner by choosing the nuisance parameters suitably.

3. TWO-STAGE PROCEDURES FOR INFERRING ABOUT THE UNIVARIATE NORMAL MEAN

In this section we suppose a population $N(\mu, \sigma^2)$ is given. This, in effect, means an unlimited sequence of i.i.d. random variables $\tilde{y}_1, \tilde{y}_2, \ldots$ all distributed as $N(\mu, \sigma^2)$ is available. Both the parameters μ and σ^2 are supposed to be unknown; we want to infer about μ.

3.1 Interval Estimation

First consider the problem of finding a confidence interval for μ with specified confidence coefficient $1 - \alpha$ $(0 < \alpha < 1)$ and span $2d$ $(0 < d < \infty)$. We start with a pilot sample size $n_0 (\geq 2)$ and a number $z > 0$ (to be determined later in terms of α, d, and n_0).

3.1.1 Stein's First Procedure

STAGE I: Take a pilot sample of size n_0: $\tilde{y}_1, \ldots, \tilde{y}_{n_0}$. Compute the mean $\bar{\tilde{y}}_0$ and the usual unbiased estimate s_0^2 of σ^2 (with $n_0 - 1 = \nu_0$ d.f.) from the pilot sample (we take the sure event $s_0^2 > 0$ for granted). Determine the smallest integer N such that $N > n_0$, $N \geq z^{-1} s_0^2$. Writing $[x]$ for the largest integer *less than* the positive number x, this means

$$N = \max\{n_0 + 1, [z^{-1} s_0^2] + 1\}. \tag{3.1}$$

Then choose $1 + N - n_0$ numbers $a_0, a_{n_0+1}, \ldots, a_N$ satisfying the equations

$$a_0 + \sum_{k=n_0+1}^{N} a_k = 1, \tag{3.2}$$

$$\frac{1}{n_0} a_0^2 + \sum_{k=n_0+1}^{N} a_k^2 = z s_0^{-2}. \tag{3.3}$$

Note that subject to (3.2) the left-hand member of (3.3) can attain any value in the range $[N^{-1}, \infty)$. Since (3.1) implies $z s_0^{-2} \geq N^{-1}$, equations (3.2)–(3.3) are solvable. We assume some convention (e.g., $a_{n_0+1} = \cdots = a_N$), which determines $a_0, a_{n_0+1}, \ldots, a_N$ as soon as s_0^2 is fixed, is followed.

STAGE II Take another sample $\tilde{y}_{n_0+1} \cdots, \tilde{y}_N$ of size $N - n_0$.

After this set up the statistic

$$l_N = a_0 \bar{\tilde{y}}_0 + \sum_{k=n_0+1}^{N} a_k \tilde{y}_k. \tag{3.4}$$

Since $\bar{\tilde{y}}_0$ and s_0^2 are independently distributed, from our construction we get that, given s_0^2, l_N is conditionally distributed as $N(\mu, z\sigma^2 s_0^{-2})$. This means $z^{-1/2}(l_N - \mu) s_0 / \sigma$ is distributed as $N(0, 1)$ conditionally given s_0^2, and hence, independently of s_0^2 as well. Since $\nu_0 s_0^2 / \sigma^2$ is distributed as a $\chi^2(\nu_0)$ (central chi square with d.f. ν_0), we conclude that the following theorem holds.

THEOREM 3.1 *With l_N defined by (3.1)–(3.4), $z^{-1/2}(l_N - \mu)$ is distributed as $t(\nu_0)$—i.e., as a Student's t with d.f. ν_0.*

If $\pm t(\nu_0, \alpha/2)$ are points in the $t(\nu_0)$-distribution containing a probability $1 - \alpha$ between them, we get

$$P\{-z^{1/2} t(\nu_0, \tfrac{1}{2}\alpha) < l_N - \mu < z^{1/2} t(\nu_0, \tfrac{1}{2}\alpha)\} = 1 - \alpha. \tag{3.5}$$

If initially z had been chosen to satisfy

$$z^{1/2} t(\nu_0, \tfrac{1}{2}\alpha) = d, \tag{3.6}$$

(3.5) would imply

$$P\{l_N - d < \mu < l_N + d\} = 1 - \alpha \tag{3.7}$$

whatever μ, σ^2. Thus $(l_N - d, l_N + d)$ would be a $(1 - \alpha)$-confidence interval for μ with predetermined span $2d$.

3.1.2 Stein's Second Procedure

The above procedure yields for μ a confidence interval with confidence coefficient exactly equal to $1 - \alpha$ and span exactly equal to $2d$. If we are prepared to replace

the phrase "exactly equal to" in one of these places by the phrase "bounded by" ("below" in the case of confidence coefficient, and "above" in the case of span), we can modify the procedure as follows to get one which is at the same time slightly more economical in sampling terms.

STAGE I After determining s_0^2 as before, find

$$N = \max\{n_0, [z^{-1}s_0^2] + 1\}. \tag{3.8}$$

STAGE II If N given by (3.8) is greater than n_0, take $N - n_0$ additional observations $\tilde{y}_{n_0+1}, \ldots, \tilde{y}_N$; if $N = n_0$, terminate sampling at the first stage.

Let $\tilde{\bar{y}}_N$ denote the overall mean of the N observations taken i.e.,

$$\begin{aligned} \tilde{\bar{y}}_N &= N^{-1}\left(n_0\bar{\tilde{y}}_0 + \sum_{k=n_0+1}^{N} \tilde{y}_k\right) \qquad \text{if } N > n_0, \\ &= \bar{\tilde{y}}_0 \qquad \text{if } N = n_0. \end{aligned} \tag{3.9}$$

Given s_0^2, N is fixed. Therefore $N^{1/2}(\tilde{\bar{y}}_N - \mu)/\sigma$ is distributed as $N(0,1)$ conditionally given s_0^2, and hence independently of s_0^2. From this we get the next theorem.

THEOREM 3.2 *y_N being defined by (3.8)–(3.9), $N^{1/2}(\tilde{\bar{y}}_N - \mu)/s_0$ is distributed as $t(\nu_0)$.*

This immediately leads to

$$P\left\{\tilde{\bar{y}}_N - N^{-1/2}s_0 t(\nu_0, \tfrac{1}{2}\alpha) < \mu < \tilde{\bar{y}}_N + N^{-1/2}s_0 t(\nu_0, \tfrac{1}{2}\alpha)\right\} = 1 - \alpha, \tag{3.10}$$

whatever μ, σ^2. Since (3.8) implies $N \geq z^{-1}s_0^2$, if z is as in (3.6), we would have

$$N^{-1/2}s_0 t(\nu_0, \tfrac{1}{2}\alpha) \leq d. \tag{3.11}$$

Thus (3.10) gives a $(1-\alpha)$-confidence interval for μ with span bounded above by $2d$. Alternatively, (3.10)–(3.11) can be construed to mean that $(\tilde{\bar{y}}_N - d, \tilde{\bar{y}}_N + d)$ represents a $2d$-span confidence interval for μ with confidence coefficient bounded below by $1 - \alpha$.

3.2 Testing

Consider the problem of testing $H^{(0)}$: $\mu = \mu^{(0)}$ against either $H^{(1)}$: $\mu > \mu^{(0)}$ or $H^{(2)}$: $\mu \neq \mu^{(0)}$. If we follow Stein's first procedure and define l_N by (3.1)–(3.4), from Theorem 3.1 we get that, under $H^{(0)}$,

$$t = z^{-1/2}(l_N - \mu^{(0)}) \tag{3.12}$$

is distributed as $t(\nu_0)$. Generally, when μ is true, $t = z^{-1/2}(l_N - \mu) + z^{-1/2}(\mu - \mu^{(0)})$ is distributed as $t(\nu_0) + z^{-1/2}(\mu - \mu^{(0)})$. Hence if we use the one-sided rule

$$\text{Reject } H^{(0)} \text{ iff } t > t(\nu_0, \alpha) \tag{3.13}$$

for testing $H^{(0)}$ against $H^{(1)}$, the size of the test is α and the power function is

$$P\left\{t(\nu_0) > t(\nu_0, \alpha) - z^{-1/2}(\mu - \mu^{(0)})\right\}, \tag{3.14}$$

which is free from σ. (Contrast this with the power function of the standard t-test based on a sample of fixed size n, which involves σ through the noncentrality parameter $n^{1/2}(\mu - \mu^{(0)})/\sigma$.) Clearly (3.14) is strictly increasing in $z^{-1/2}(\mu - \mu^{(0)})$. If we want the power to be greater than a specified value β $(\alpha < \beta < 1)$ for $\mu > \mu^{(0)} + \delta$, where $\delta > 0$ is given, we have only to choose z so as to satisfy

$$P\left\{t(\nu_0) > t(\nu_0, \alpha) - z^{-1/2}\delta\right\} = \beta. \tag{3.15}$$

Further, rule (3.13) also clearly gives a size α test for $\mu \leq \mu^{(0)}$ against $H^{(1)}$.

To test $H^{(0)}$ against $H^{(2)}$ at level α, we can use the two-sided rule

$$\text{Reject } H^{(0)} \text{ iff } |t| > t(\nu_0\, \tfrac{1}{2}\alpha) \tag{3.16}$$

Since the $t(\nu_0)$-distribution is symmetric, bell shaped, it readily follows that the power function of (3.16) is strictly increasing in $z^{-1/2}|\mu - \mu^{(0)}|$ and that z can be chosen so that the power exceeds any prescribed β $(\alpha < \beta < 1)$ for $|\mu - \mu^{(0)}| > \delta$.

Just as for interval estimation, we can improve the above tests by following Stein's second procedure. If $\tilde{\bar{y}}_N$ is given by (3.8)–(3.9), we define

$$t^* = N^{1/2}\left(\tilde{\bar{y}}_N - \mu^{(0)}\right)/s_0. \tag{3.17}$$

Since by Theorem 3.2, t^* follows a $t(\nu_0)$-distribution under $H^{(0)}$, we can formulate one-sided or two-sided tests for $H^{(0)}$ against $H^{(1)}$ or $H^{(2)}$ on its basis. The power functions of these tests, though involving σ, would for relevant alternatives always dominate those of the corresponding test based on t (n_0 and z being same). Thus the power of the one-sided test based on t^* against $H^{(1)}$ would be, for $\mu > \mu^{(0)}$,

$$\begin{aligned} &P\left\{t^* > t(\nu_0, \alpha) \mid \mu, \sigma\right\} \\ &\quad = P\left\{N^{1/2}s_0^{-1}(\tilde{\bar{y}} - \mu) > t(\nu_0, \alpha) - N^{1/2}s_0^{-1}(\mu - \mu^{(0)}) \mid \mu, \sigma\right\} \\ &\quad > P\left\{N^{1/2}s_0^{-1}(\tilde{\bar{y}}_N - \mu) > t(\nu_0, \alpha) - z^{-1/2}(\mu - \mu^{(0)}) \mid \mu, \sigma\right\} \\ &\qquad \text{(since (3.8) means } N \geq z^{-1}s_0^2\text{)} \\ &\quad = P\left\{t(\nu_0) > t(\nu_0, \alpha) - z^{-1/2}(\mu - \mu^{(0)})\right\}, \end{aligned} \tag{3.18}$$

the last expression being identical with (3.14). It follows that if z is determined from (3.15), (3.18) would be greater than β for $\mu > \mu^{(0)} + \delta$ irrespective of σ. Similar observations hold for the test based on t^* against the two-sided alternative $H^{(2)}$.

3.3 Point Estimation

The two-stage procedures of Stein can also be used to obtain for a point estimate with expected loss equal to or bounded by a specified value for loss functions of a

standard form (Lehmann 1950). Thus suppose the loss in estimating μ by a has the form $W(|a-\mu|)$, where $W(x)$ is a given function nondecreasing in $[0,\infty)$ with $W(0)=0$, $W(\infty)\le\infty$, such that $EW(c|t(\nu_0)|)$ exists for the chosen ν_0, for any $c>0$. (In particular, the absolute error loss with $W(x)=x$ and the squared error loss with $W(x)=x^2$ are of this type, provided $\nu_0\ge 2$ and $\nu_0\ge 3$, respectively.) Then, if we use Stein's first procedure and take l_N as our estimate, by Theorem 3.1 the expected loss would be

$$E_{\nu,\sigma^2}W(|l_N-\mu|) = EW(z^{1/2}|t(\nu_0)|) = z^{-1/2}\int_{-\infty}^{\infty} W(|x|)f_{\nu_0}(z^{1/2}x)\,dx, \tag{3.19}$$

where $f_{\nu_0}(\cdot)$ stands for the density function of $t(\nu_0)$. From our assumptions regarding $W(\cdot)$, it follows that (3.19) is a continuous nondecreasing function of $z\in(0,\infty)$ which approaches 0 and $W(\infty)$ as z tends to 0 and ∞, respectively. Therefore given any number w, $0<w<W(\infty)$, we can determine z so that

$$E_{\nu,\sigma^2}W(|l_N-\mu|)=w. \tag{3.20}$$

Further, if we use Stein's second procedure with the same n_0 and z and take $\tilde{\bar{y}}_N$ given by (3.8)–(3.9) as our point estimate, by Theorem 3.2 the expected loss would be

$$E_{\mu,\sigma^2}W(|\tilde{\bar{y}}_N-\mu|) \le E_{\mu,\sigma^2}W(z^{1/2}N^{1/2}s_0^{-1}|\tilde{\bar{y}}_N-\mu|) = EW(z^{1/2}|t(\nu_0)|)=w. \tag{3.21}$$

3.4 Average Sample Number of the Procedures

From (3.1) and (3.8) it is straightforward to derive expressions for the average sample number (ASN) of Stein's two procedures. Thus considering the second procedure, since $\nu_0 s_0^2/\sigma^2$ is distributed as a $\chi^2(\nu_0)$, by (3.8),

$$\begin{aligned} E_{\sigma^2}(N) &= n_0 P\{z^{-1}s_0^2\le n_0 \mid \sigma^2\} + \sum_{n=n_0}^{\infty}(n+1)P\{n<z^{-1}s_0^2\le n+1\mid\sigma^2\} \\ &= n_0 P\{\chi^2(\nu_0)\le \nu_0 z\sigma^{-2}n_0\} + \sum_{n=n_0}^{\infty}(n+1)P\{\nu_0 z\sigma^{-2}n<\chi^2(\nu_0)\le\nu_0 z\sigma^{-2}(n+1)\} \end{aligned} \tag{3.22}$$

Further, since whenever $n_0<z^{-1}s_0^2$, we have $z^{-1}s_0^2\le N\le z^{-1}s_0^2+1$, writing

$$b(\sigma^2)=n_0P\{\chi^2(\nu_0)\le\nu_0z\sigma^{-2}n_0\}+z^{-1}\sigma^2P\{\chi^2_{(\nu_0+2)}>\nu_0z\sigma^{-2}n_0\}, \tag{3.23}$$

we get the sharp bounds

$$b(\sigma^2)<E_{\sigma^2}(N)<b(\sigma^2)+P\{\chi^2(\nu_0)>\nu_0z\sigma^{-2}n_0\}. \tag{3.24}$$

When n_0 and $z\sigma^{-2}$ are both small, we get

$$E_{\sigma^2}(N)\simeq z^{-1}\sigma^2, \tag{3.25}$$

which is intuitively evident from (3.8). Under the same conditions (3.25) holds for Stein's first procedure also.

From these expressions it is apparent that, for fixed z, $E_{\sigma^2}(N)$ increases as σ^2 increases and, for fixed σ^2, decreases as z decreases. These conclusions agree with our intuition: σ^2 larger means the population is more variable so that we have to pay a higher price to achieve the same level of performance for our inference procedure; z smaller means our requirement on the performance of the procedure is more stringent and therefore, the cost is also correspondingly higher.

The above detailed discussion of Stein's procedures for a normal mean shows how these achieve what we have called *comprehensive studentization* in Section 1. The performance of the inference procedure is made entirely free of the nuisance parameter σ by transferring the influence of σ to the sampling behavior of the ultimate sample size N. The idea has been employed in more complex setups to develop inference procedures for linear functions of parameters of location and regression in the presence of one or more nuisance parameters. We discuss some of these in the following sections.

4. TWO-STAGE PROCEDURES FOR THE REPLICABLE LINEAR REGRESSION AITKEN SETUP

Suppose an unlimited sequence of observable i.i.d. random p-vectors $\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \ldots$, each following $N_p(X\boldsymbol{\beta},\sigma^2\Lambda)$, is given. Here

- $X^{p\times q}$ is a known matrix whose rank (without loss of generality) is assumed to be q, $1\le q\le p$.
- $\Lambda^{p\times p}$ is a known positive definite (p.d.) matrix.
- $\boldsymbol{\beta}^{q\times 1}$ is an unknown vector of regression parameters, $\boldsymbol{\beta}\in\mathcal{R}^q$.
- σ^2 is an unknown positive scalar, $\sigma^2\in(0,\infty)$

We want to infer about r ($1\le r\le q$) linear functions of the regression parameters represented by $\underline{H}\boldsymbol{\beta}$, where $\underline{H}^{r\times q}$ is a given matrix of rank r (assumed without loss of generality). Clearly this is a linear regression setup in Aitken's form; i.e., ratios of the variances and the pairwise correlation coefficients are all assumed to be known. It is replicable in the sense that the experiment involving the observation of the entire set of p variables can be replicated independently any number of times.

A two-stage procedure for a canonically reduced (in terms of both X and Λ) version of this setup was developed by Stein (1945). We give here the procedure for the setup in its unreduced form. Incidentally, writing $E(\tilde{\mathbf{y}}_k)=\boldsymbol{\mu}$, here we are assuming that $\boldsymbol{\mu}$ varies in the column space of X. When $\boldsymbol{\mu}$ varies in $\mathcal{R}^p$ corresponds to the case $q=p$, $X=I_p$, $\boldsymbol{\mu}=\boldsymbol{\beta}$.

We start with an integer n_0 such that $\nu_0=n_0p-q>1$ and take a pilot sample $\tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{n_0}$ of size n_0. Let s_0^2 be the standard unbiased estimate of σ^2 based on

the ν_0 d.f. available for the pilot sample. This means

$$\nu_0 s_0^2 = \sum_{k=1}^{n_0} (\tilde{\mathbf{y}}_k - X\hat{\boldsymbol{\beta}}_0)\,\Lambda^{-1}(\tilde{\mathbf{y}}_k - X\hat{\boldsymbol{\beta}}_0), \tag{4.1}$$

where

$$\hat{\boldsymbol{\beta}}_0 = (X\Lambda^{-1}X)^{-1}X'\Lambda^{-1}\bar{\tilde{\mathbf{y}}}_0, \bar{\tilde{\mathbf{y}}}_0 = n_0^{-1}\sum_{k=1}^{n_0}\tilde{\mathbf{y}}_k. \tag{4.2}$$

Taking any $z > 0$, we then define N as in (3.1) and find $1 + N - n_0$ numbers a_0, $a_{n_0+1}, \cdots, a_N$ satisfying (3.2)–(3.3), with s_0^2 now given by (4.1)–(4.2). Let

$$\mathbf{l}_N^{p\times 1} = a_0\bar{\tilde{\mathbf{y}}}_0 + \sum_{k=n_0+1}^{N} a_k\tilde{\mathbf{y}}_k. \tag{4.3}$$

Given s_0^2, $\mathbf{l}_N$ follows the distribution $N_p(X\boldsymbol{\beta}, zs_0^{-2}\sigma^2\Lambda)$. Now let $\hat{\boldsymbol{\beta}}_N$ denote the least squares estimate of $\boldsymbol{\beta}$ based on $\mathbf{l}_N$ obtained by minimizing $(\mathbf{l}_N - X\boldsymbol{\beta})'\Lambda^{-1}(\mathbf{l}_N - X\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$. This means

$$\boldsymbol{\beta}_N = (X'\Lambda^{-1}X)^{-1}X'\Lambda^{-1}\mathbf{l}_N. \tag{4.4}$$

The corresponding estimate of $\underline{H}\boldsymbol{\beta}$ would be $\underline{H}\hat{\boldsymbol{\beta}}_N$. From (4.4) we get that, given s_0^2, $\underline{H}\hat{\boldsymbol{\beta}}_N$ is distributed as $N_r(\underline{H}\boldsymbol{\beta}, zs_0^{-2}\sigma^2\underline{H}(X'\Lambda^{-1}X)^{-1}\underline{H}')$. This result can be utilized to formulate a confidence region and a test for $\underline{H}\boldsymbol{\beta}$.

Considering the confidence region first, we write

$$Q = z^{-1}r^{-1}(\underline{H}\hat{\boldsymbol{\beta}}_N - \underline{H}\boldsymbol{\beta})(\underline{H}(X'\Lambda^{-1}X)^{-1}\underline{H}')^{-1}(\underline{H}\hat{\boldsymbol{\beta}}_N - \underline{H}\boldsymbol{\beta}). \tag{4.5}$$

The above result implies rs_0^2Q/σ^2 is distributed as a $\chi^2(r)$ conditionally given s_0^2, and hence, independently of s_0^2. Since $\nu_0 s_0^2/\sigma^2$ follows a $\chi^2(\nu_0)$-distribution, this means Q is distributed as $F(r, \nu_0)$ (central F with d.f. r, ν_0). Therefore if $F(r, \nu_0; \alpha)$ is the upper α-point of this distribution, we have

$$(\underline{H}\boldsymbol{\beta} - \underline{H}\hat{\boldsymbol{\beta}}_N)'(\underline{H}(X'\Lambda^{-1}X)^{-1}\underline{H}')^{-1}(\underline{H}\boldsymbol{\beta} - \underline{H}\hat{\boldsymbol{\beta}}_N) \le zrF(r, \nu_0; \alpha) \tag{4.6}$$

as a $(1 - \alpha)$-confidence region for $\underline{H}\boldsymbol{\beta}$. The shape of this ellipsoidal region is determined by Λ, X, and $\underline{H}$, which are all known; the span (or volume) depends on z, which can be chosen suitably to give the region any desired maximum diameter or volume.

If our object is to test a hypothesis $H^{(0)}$: $\underline{H}\boldsymbol{\beta} = \boldsymbol{\psi}^{(0)}$, where $\boldsymbol{\psi}^{(0)}$ is a given r-vector, against all alternatives, we can taken

$$Q^{(0)} = z^{-1}r^{-1}(\underline{H}\hat{\boldsymbol{\beta}}_N - \boldsymbol{\psi}^{(0)})'(\underline{H}(X'\Lambda^{-1}X)^{-1}\underline{H}')^{-1}(\underline{H}\hat{\boldsymbol{\beta}}_N - \boldsymbol{\psi}^{(0)}) \tag{4.7}$$

as our test statistic. From the above, it follows that under $H^{(0)}$, $Q^{(0)}$ follows an $F(r, \nu_0)$-distribution. Thus we can set up a right-tailed test based on $Q^{(0)}$ for

$H^{(0)}$ with specified size α. Reasoning as earlier, we get that, given s_0^2, $rs_0^2Q^{(0)}/\sigma^2$ is conditionally distributed as a $\chi'^2(r \mid v)$ (noncentral chi square with d.f.r. and noncentrality parameter v) with $v = z^{-1}s_0^2\Delta/\sigma^2$, where

$$\Delta = (\underline{H}\boldsymbol{\beta} - \boldsymbol{\psi}^{(0)})'(\underline{H}(X'\Lambda^{-1}X)^{-1}\underline{H}')^{-1}(\underline{H}\boldsymbol{\beta} - \boldsymbol{\psi}^{(0)}). \tag{4.8}$$

If we now integrate out s_0^2/σ^2, which is distributed as $\nu_0^{-1}\chi^2(\nu_0)$, we get that the unconditional distribution of $Q^{(0)}$ involves only $z^{-1}\Delta$ (besides r and ν_0) and is free from σ^2. (For the form of this distribution see Chatterjee (1962), Stein (1945), or Ruben (1961).) The power of the right-tailed test based on $Q^{(0)}$ is strictly increasing in $z^{-1}\Delta$. We can choose z suitably to ensure that the power is never less than a specified level for values of Δ exceeding a given value.

Just as in the case of the normal mean, we can modify our procedure by replacing (3.1) by (3.8) to improve upon the above confidence region and test. For the modified procedure defining $\bar{\tilde{\mathbf{y}}}_N$ as in (3.9), we find

$$\hat{\boldsymbol{\beta}}_N^* = (X'\Lambda^{-1}X)^{-1}X'\Lambda^{-1}\bar{\tilde{\mathbf{y}}}_N. \tag{4.9}$$

Arguing as before, it follows that, given s_0^2, $\underline{H}\hat{\boldsymbol{\beta}}_N^*$ is distributed as $N_r(\underline{H}\boldsymbol{\beta}, N^{-1}\sigma^2\underline{H}(X'\Lambda^{-1}X)^{-1}\underline{H}')$ and, hence,

$$Nr^{-1}s_0^{-2}(\underline{H}\hat{\boldsymbol{\beta}}_N^* - \underline{H}\boldsymbol{\beta})'(\underline{H}(X'\Lambda^{-1}X)^{-1}\underline{H}')^{-1}(\underline{H}\hat{\boldsymbol{\beta}}_N^* - \underline{H}\boldsymbol{\beta}) \tag{4.10}$$

is distributed as $F(r, \nu_0)$. We can get a confidence region for $\underline{H}\boldsymbol{\beta}$ which has the same form as (4.6) with the right-hand member replaced by $rs_0^2F(r, \nu_0; \alpha)/N$. Since $N \ge z^{-1}s_0^2$, this modified confidence region is smaller than the confidence region (4.6) in the sense that when the two are placed at the same center by translation, the former always lies inside the latter. Similarly we can formulate a test for $H^{(0)}$: $\underline{H}\boldsymbol{\beta} = \boldsymbol{\psi}^{(0)}$ in terms of the statistic Q^* which would be obtained by replacing $\underline{H}\boldsymbol{\beta}$ in (4.10) by $\boldsymbol{\psi}^{(0)}$. Under an alternative, given s_0^2, $rs_0^2Q^*/\sigma^2$ is conditionally distributed as $\chi'^2(r \mid v^*)$, where $v^* = N\Delta/\sigma^2$. Since $N \ge z^{-1}s_0^2$, this implies the power of the test based on Q^* dominates the power of that based on $Q^{(0)}$.

Just as in Section 3, we can use $\underline{H}\hat{\boldsymbol{\beta}}_N$ or $\underline{H}\hat{\boldsymbol{\beta}}_N^*$ as a point estimate of $\underline{H}\boldsymbol{\beta}$. If the loss for an estimate $\mathbf{a}^{r\times 1}$ of $\underline{H}\boldsymbol{\beta} = \boldsymbol{\psi}$ is a known increasing function of the "squared distance" $(\mathbf{a} - \boldsymbol{\psi})'(\underline{H}(X'\Lambda^{-1}X)^{-1}\underline{H}')^{-1}(\mathbf{a} - \boldsymbol{\psi})$, we can control the expected loss by adjusting z.

Obviously, the ASN of the procedures of this section are identical to those of Section 3, with ν_0 now given by $n_0p - q$.

5. TWO-STAGE PROCEDURES FOR INFERRING ABOUT THE MULTINORMAL MEAN VECTOR

In this section we assume that a sequence of observable i.i.d. random p-vectors $\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \ldots$ following a common distribution $N_p(\boldsymbol{\mu}, \Sigma)$ is given. Our object is to infer about the unknown mean vector $\boldsymbol{\mu}$; the unknown dispersion Σ can be any p.d. matrix. We consider two two-stage procedures here: the first gives for $\boldsymbol{\mu}$ an

ellipsoidal confidence region with span (maximum diameter) subject to a given upper bound, the second gives an ellipsoidal confidence region with a previously fixed contour and also a test with power free from Σ.

5.1 Bounded-Span Confidence Region

We first take a pilot sample $\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{n_0}$ of size n_0 $(\geq p+1)$ and obtain from it the standard unbiased estimate S_0 of Σ based on $\nu_0 = n_0 - 1$ d.f. Specifically,

$$\nu_0 S_0 = \sum_{k=1}^{n_0} (\tilde{\mathbf{y}}_k - \bar{\tilde{\mathbf{y}}}_0)(\tilde{\mathbf{y}}_k - \bar{\tilde{\mathbf{y}}}_0)', \qquad \bar{\tilde{\mathbf{y}}}_0 = n_0^{-1} \sum_{k=1}^{n_0} \tilde{\mathbf{y}}_k. \tag{5.1}$$

Let $\lambda_M(S_0)$ denote the largest characteristic root of S_0. We find (cf. (3.8))

$$N = \max\{n_0, [z^{-1}\lambda_m(S_0)] + 1\}, \tag{5.2}$$

$z > 0$ being, as before, an arbitrarily fixed number. If $N > n_0$, we take $N - n_0$ additional observations $\tilde{\mathbf{y}}_{n_0+1}, \ldots, \tilde{\mathbf{y}}_N$. Let $\bar{\tilde{\mathbf{y}}}_N$ $(= \bar{\tilde{\mathbf{y}}}_0$ if $N = n_0)$ denote the mean vector of all the N observations. Arguing as in the case of Theorem 3.2, we get that $N^{1/2}(\bar{\tilde{\mathbf{y}}}_N - \boldsymbol{\mu})$ is distributed as $N_p(0, \Sigma)$, conditionally given S_0, and hence, independently of S_0. Since $\nu_0 S_0$ follows a Wishart distribution with parent dispersion Σ and d.f ν_0, we get that

$$N(\bar{\tilde{\mathbf{y}}}_N - \boldsymbol{\mu})' S_0^{-1} (\bar{\tilde{\mathbf{y}}}_N - \boldsymbol{\mu}) \tag{5.3}$$

follows a central Hotelling T^2-distribution with d.f $(p, \nu_0 - p + 1)$. If $T^2(p, \nu_0 - p + 1; \alpha) = T^2(\alpha)$ denotes the upper α-point of this distribution,

$$(\boldsymbol{\mu} - \bar{\tilde{\mathbf{y}}}_N)' S_0^{-1} (\boldsymbol{\mu} - \bar{\tilde{\mathbf{y}}}_N) \leq N^{-1} T^2(\alpha) \tag{5.4}$$

is a $(1-\alpha)$-confidence region for $\boldsymbol{\mu}$. The contour of this ellipsoidal region depends on S_0, but by (5.2) its maximum diameter is

$$2\{N^{-1}T^2(\alpha)\lambda_M(S_0)\}^{1/2} \leq 2\{zT^2(\alpha)\}^{1/2}. \tag{5.5}$$

The upper bound on the right of (5.5) can be given any desired value by choosing z.

The above confidence region was proposed by Healy (1956). Naturally a test for the hypothesis $\boldsymbol{\mu} = \boldsymbol{\mu}^{(0)}$ can be based on the statistic (5.3) (with $\boldsymbol{\mu}$ replaced by $\boldsymbol{\mu}^{(0)}$). However, the power of that test would involve Σ, and it does not seem possible to give a useful Σ-free lower bound to that power. In the next subsection we describe an alternative two-stage procedure which yields a confidence region with a predetermined ellipsoidal contour as well as a test with Σ-free power.

5.2 Fixed-Contour Confidence Region and Related Tests

We want a confidence region for $\boldsymbol{\mu}$ which has a stipulated confidence coefficient $1-\alpha$ and the form

$$(\boldsymbol{\mu} - \mathbf{c})' \Omega (\boldsymbol{\mu} - \mathbf{c}) \leq d^2, \tag{5.6}$$

where Ω is a given p.d. matrix and $d > 0$ is a given number. The center $\mathbf{c}$ is unspecified; naturally $\mathbf{c}$ would depend on the sample and would vary randomly so that, when $\boldsymbol{\mu}$ is true, (5.6) holds with probability $1-\alpha$.

We start with an integer n_0 $(\geq p+1)$ and an arbitrary number $z > 0$. Taking a pilot sample size n_0, we determine S_0 as in (5.1). Let

$$N = \max\{n_0 + p^2, [z^{-1} \operatorname{tr}(\Omega S_0)] + 1\}. \tag{5.7}$$

We take $N - n_0$ additional observations $\tilde{\mathbf{y}}_{n_0+1}, \ldots, \tilde{\mathbf{y}}_N$ and write

$$\tilde{Y}_{0N}^{p \times (1+N-n_0)} = (\tilde{\mathbf{y}}_0, \tilde{\mathbf{y}}_{n_0+1}, \ldots, \tilde{\mathbf{y}}_N). \tag{5.8}$$

Denoting

$$\mathbf{1}' = (1, 1, \ldots, 1)^{1 \times (1+N-n_0)}, \qquad D = \operatorname{diag}(n_0^{-1}, 1, \ldots, 1)^{(1+N-n_0) \times (1+N-n_0)}$$

and using $\otimes$ for Kronecker product, we next determine p $p \times (1+N-n_0)$ matrices:

$$A_i = (\mathbf{a}_{i1}, \mathbf{a}_{i2}, \ldots, \mathbf{a}_{ip})', \qquad i = 1, 2, \ldots, p, \tag{5.9}$$

such that if

$$A' = (A_1', A_2', \ldots, A_p'), \tag{5.10}$$

then

$$\begin{aligned} &\text{(i)} \quad (A_1\mathbf{1}, A_2\mathbf{1}, \ldots, A_p\mathbf{1}) = I_p, \\ &\text{(ii)} \quad ADA' = z\Omega^{-1} \otimes S_0^{-1}. \end{aligned} \tag{5.11}$$

If N is given by (5.7), A_1, A_2, ..., A_p subject to (5.11) can be determined. This can be seen by applying the following lemma repeatedly to determine the p^2 rows of A (see Chatterjee 1959a).

LEMMA 5.1 *Given a p.d. matrix $G^{m \times m}$, a matrix $C^{m \times r}$ such that $C'C$ is nonsingular, a vector $\mathbf{b}^{r \times 1}$ and a number $f^2 > 0$, a vector $\mathbf{a}^{m \times 1}$ subject to*

$$\mathbf{a}'C = b', \qquad \mathbf{a}'G\mathbf{a} = f^2$$

can be determined only if $m \geq r$ and iff

$$\left|\begin{array}{cc} C'G^{-1}C & \mathbf{b} \\ \mathbf{b}' & f^2 \end{array}\right| \begin{cases} = 0 & \text{for } m = r, \\ \geq 0 & \text{for } m \geq r+1. \end{cases}$$

(For a proof see the appendix in Chatterjee 1962a.)

We suppose some definite rule is followed to determine $A_1, \ldots, A_p$ so that these become fixed as soon as S_0 is fixed.

We next set up

$$l_{iN} = \operatorname{tr}(A_i \tilde{Y}'_{0N}), \qquad i = 1, \ldots, p,$$
$$\mathbf{l}_N = (l_{1N}, \ldots, l_{pN})'. \tag{5.12}$$

Given S_0, N becomes fixed, and by (5.11), conditionally $\mathbf{l}_N$ is distributed as $N_p(\boldsymbol{\mu}, zv\Omega^{-1})$, where

$$v = \operatorname{tr}(S_0^{-1}\Sigma). \tag{5.13}$$

Writing now

$$u = z^{-1}(\mathbf{l}_N - \boldsymbol{\mu})'\Omega(\mathbf{l}_N - \boldsymbol{\mu}), \tag{5.14}$$

it follows that $v^{-1}u$ is distributed as a $\chi^2(p)$ conditionally given S_0 and, hence, independently of S_0. Now from (5.1) and (5.13) it is clear that v is invariant under the transformation $S_0 \to BS_0B'$, $\Sigma \to B\Sigma B'$, where $B^{p\times p}$ is any nonsingular matrix. Hence it is immediate that the distribution of v is free from Σ. Denoting the density of v, which depends only on p and ν_0 by $g_{p,\nu_0}(v)$ and the density of $\chi^2(p)$ by $f_{\chi^2(p)}(\cdot)$, we conclude the following.

THEOREM 5.1 *The distribution of u defined by (5.7), (5.8), (5.11), (5.12), and (5.14) depends only on p and ν_0, and its density is given by*

$$\int_0^\infty v^{-1} f_{\chi^2(p)}(v^{-1}u) g_{p,\nu_0}(v)\, dv, \qquad 0 < u < \infty, \tag{5.15}$$

The above theorem can be utilized to construct a confidence region and a test for $\boldsymbol{\mu}$. Thus if $u(p, \nu_0; \alpha) = u(\alpha)$ is the upper α-point of the distribution (5.15), we would have

$$P\{(\mathbf{l}_N - \boldsymbol{\mu})'\Omega(\mathbf{l}_N - \boldsymbol{\mu}) \leq zu(\alpha)\} = 1 - \alpha,$$

whatever $\boldsymbol{\mu}, \Sigma$. Hence if z is chosen to satisfy

$$zu(\alpha) = d^2, \tag{5.16}$$

we would get

$$(\boldsymbol{\mu} - \mathbf{l}_N)'\Omega(\boldsymbol{\mu} - \mathbf{l}_N) \leq d^2 \tag{5.17}$$

as a $(1-\alpha)$-confidence region of the form (5.6).

Similarly to test $H^{(0)}$: $\boldsymbol{\mu} = \boldsymbol{\mu}^{(0)}$ against all alternatives, we may use the rule

$$\text{Reject } H_0 \text{ iff } u^{(0)} = z^{-1}(\mathbf{l}_N - \boldsymbol{\mu}^{(0)})'\Omega(\mathbf{l}_N - \boldsymbol{\mu}^{(0)}) > u(\alpha). \tag{5.18}$$

By Theorem 5.1 the test would have exact size α. Also, arguing as before, when $H^{(0)}$ is not true, given S_0, and, hence, given v, $v^{-1}u^{(0)}$ is conditionally distributed as $\chi'^2(p \mid v^{-1}z^{-1}\Delta)$, where

$$\Delta = (\boldsymbol{\mu} - \boldsymbol{\mu}^{(0)})'\Omega(\boldsymbol{\mu} - \boldsymbol{\mu}^{(0)}). \tag{5.19}$$

Hence, denoting the density of $\chi'^2(p \mid \Delta^*)$ by $f_{\chi'^2(p)}(\cdot \mid \Delta^*)$, we find that the power function of test (5.18) is

$$\int_0^\infty \left[\int_{v^{-1}u(\alpha)}^\infty f_{\chi'^2(p)}(x \mid v^{-1}z^{-1}\Delta)\, dx\right] g_{p,\nu_0}(v)\, dv. \tag{5.20}$$

This is free from Σ and depends on the alternative only through the "distance" Δ given by (5.19). The inner integral in (5.20), and hence (5.20) itself, is monotonic in $z^{-1}\Delta$, (5.20) increasing from α to 1 as $z^{-1}\Delta$ varies from 0 to ∞. Hence we can choose z so as to ensure that the power attains a given value $\beta \in (\alpha, 1)$ at a specified value Δ' of Δ and exceeds β for points $\boldsymbol{\mu}$ for which Δ is more than Δ'.

To set up the confidence region (5.16)–(5.17), we need tables of the null distribution of u given by (5.15); to use test (5.18) to realize a stipulated power requirement, we need, in addition, tables of the power function (5.20). One approach for reducing (5.15) and (5.20) is to express v as $\sum_1^p \lambda_i^{-1}$, where the λ_i's are the roots of $|S_0 - \lambda\Sigma| = 0$, and to utilize the known form of the joint distribution of the λ_i's. For $p = 2$, in this way the tail area of (5.15) has been explicitly worked out and tabulated, and (5.20) has been reduced to a workable form. For general p, although integral expressions for the tail area of (5.15) and the power function (5.20) are available, it is troublesome to work these out explicitly. However, an asymptotic expansion for the tail area of (5.15) in powers of ν_0^{-1} has been obtained. (Note that for large ν_0, $v \simeq p$, so u is approximately distributed as $p\chi^2(p)$ and $p\chi'^2(p \mid p^{-1}z^{-1}\Delta)$, respectively, in the null and nonnull cases.) These results are contained in Chatterjee (1959a, 1960).

The procedure described above can be considered as the multivariate counterpart of Stein's first procedure for the normal mean considered in Section 3. Just as in the univariate case, it is possible to improve the procedure. The modified procedure giving an improved confidence region is particularly simple and similar to that considered in Section 5.1. After obtaining S_0 from a pilot sample as before, we work out $\lambda_M(\Omega S_0)$ and determine

$$N = \max\{n_0, [d^{-2}T^2(\alpha)\lambda_M(\Omega S_0)] + 1\}, \tag{5.21}$$

where $T^2(\alpha)$ is as in (5.4). Let $\tilde{\mathbf{y}}_N$ be the mean vector based on N observations ($N - n_0$ additional observations have to be taken, if required.) Then arguing as in

Section 5.1,

$$(\mu - \tilde{\bar{\mathbf{y}}}_N)' S_0^{-1} (\mu - \tilde{\bar{\mathbf{y}}}_N) \le N^{-1} T^2(\alpha) \tag{5.22}$$

is a $(1-\alpha)$-confidence region for μ. Using the fact that $v = \operatorname{tr}(S_0^{-1}\Sigma)$, we can show that the random variable u following the distribution (5.15) is stochastically larger than $T^2(p, \nu_0 - p + 1)$, so $T^2(\alpha) < u(\alpha)$. Since $\lambda_M(\Omega S_0) < \operatorname{tr}(\Omega S_0)$, we get that, under (5.16), N given by (5.7) (for the same pilot sample) is larger than N given by (5.21). Further, under (5.21), $N \ge d^{-2} T^2(\alpha) \lambda_M(\Omega S_0)$, which implies that region (5.22) lies completely within (5.17) when placed at the same center. In other words, N and $\tilde{\bar{\mathbf{y}}}_N$ being as in (5.22),

$$(\mu - \tilde{\bar{\mathbf{y}}}_N)' \Omega (\mu - \tilde{\bar{\mathbf{y}}}_N) \le d^2 \tag{5.23}$$

represents a confidence region with confidence coefficient at least $1 - \alpha$.

In the case of testing, however, it is not known whether the test corresponding to (5.22) is uniformly more powerful than the comparable test (5.18). A test more powerful and slightly more economical than (5.18) can be constructed here by modifying the procedure leading to (5.18) as follows.

After S_0 from a pilot sample of size n_0 is found, determine

$$N = \max \{ n_0 + p^2 - 1, [z^{-1} \operatorname{tr}(\Omega S_0)] + 1 \}. \tag{5.24}$$

For this N, let $\tilde{Y}_{0N}$ be as in (5.8) and let A_i^*, $i = 1, \ldots, p$, be defined as in (5.9)–(5.11), with z replaced by $N^{-1} \operatorname{tr}(\Omega S_0)$. (It can be shown by repeated application of Lemma 5.1 that for N given by (5.24) such A_i^*, $i = 1, \ldots, p$, can always be found.) Now let $\mathbf{l}_N^*$ be defined as in (5.12), with A_i replaced by A_i^*. Then it can be shown that the test

$$\text{Reject } H_0 \text{ iff } N \{\operatorname{tr}(\Omega S_0)\}^{-1} (\mathbf{l}_N^* - \mu^{(0)})' \Omega (\mathbf{l}_N^* - \mu^{(0)}) > u(\alpha) \tag{5.25}$$

has size α and power uniformly more than test (5.18). The improved test, however, does not have the simplicity of the improved confidence region (5.23).

The two-stage procedures corresponding to (5.17) and (5.22) can be used to get $\mathbf{l}_N$ and $\tilde{\bar{\mathbf{y}}}_N$, respectively, as point estimates of μ. If the loss resulting from an estimate $\mathbf{a}$ is of the form $W((\mathbf{a} - \mu)' \Omega (\mathbf{a} - \mu))$, where $W(\cdot)$ is a known increasing function, we can adjust z to make the expected loss equal to or bounded by any admissible value.

Further details of the above procedures are contained in Chatterjee (1959a,b).

6. TWO-STAGE PROCEDURES FOR THE GENERAL LINEAR REGRESSION SETUP: NONSTOCHASTIC AND STOCHASTIC PREDICTORS

We now consider a genuine linear regression setup in which the successive observations on the predicted variable correspond to possibly different levels of the

predictors. First suppose we have a sequence of independent normal observations $\tilde{y}_1, \tilde{y}_2, \ldots$ on the predictand with a common unknown variance σ^2, $\tilde{y}_k$ corresponding to the levels $(x_{1k}, \ldots, x_{qk})' = \mathbf{x}_k$ of q nonstochastic predictors. The sequence $\mathbf{x}_k$, $k = 1, 2, \ldots$, is unlimited. It is assumed that any number of terms of this sequence can be determined and the corresponding $\tilde{y}_k$'s observed. It is known that

$$E(\tilde{y}_k) = \beta' \mathbf{x}_k, \qquad k = 1, 2, \ldots, \tag{6.1}$$

where $\beta = (\beta_1, \ldots, \beta_q)'$ are unknown regression coefficients. Writing

$$X_n^{q \times n} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \tag{6.2}$$

and using $\lambda_m(\cdot)$ to denote the smallest characteristic root, we assume

$$\lim_{n \to \infty} \lambda_m(X_n X_n') = \infty. \tag{6.3}$$

(Note that $\lambda_m(X_n X_n')$ is nondecreasing in n.) Equation (6.3) implies that as n increases, X_n effectively "grows in size"; thus if, after a finite number of columns, all the remaining columns of X_n belong to a $(q-1)$-or lower-dimensional linear subspace of $\mathcal{R}^q$, equation (6.3) will not hold. In particular, by (6.3) for sufficiently large n, $\beta_1, \ldots, \beta_q$ are estimable on the basis of $\tilde{y}_1, \ldots, \tilde{y}_n$. In the following we consider inference about the full set of regression parameters $\beta_1, \ldots, \beta_q$; however, inference about a set of linear functions of the β_i's can be dealt with similarly by a straightforward modification of the procedure.

We start with an integer $n_0 \ge q + 1$ such that $\operatorname{rank}(X_{n_0}) = q$, an arbitrarily fixed number $z > 0$, and an arbitrary positive definite matrix $\Omega^{q \times q}$. Observing $\tilde{Y}_{n_0}^{n \times 1} = (\tilde{y}_1, \ldots, \tilde{y}_{n_0})'$ first and writing $\nu_0 = n_0 - q$, we find

$$\hat{\beta}_0 = (X_n X_n')^{-1} (X_n \tilde{Y}_{n_0}), \qquad \nu_0 s_0^2 = \tilde{Y}_n' \tilde{Y}_{n_0} - \hat{\beta}_0' (X_0 \tilde{Y}_{n_0}). \tag{6.4}$$

We then determine

$$N = \min \{ n : n \ge n_0 + q, (X_n X_n') - z^{-1} s_0^2 \Omega \text{ is p.d.} \}, \tag{6.5}$$

Since $z^{-1} s_0^2 \Omega$ becomes fixed as soon as s_0^2 is given, (6.3) implies that a finite N defined by (6.5) is determined by s_0^2. We next make $N - n_0$ additional observations $\tilde{y}_k$, $k = n_0 + 1, \ldots, N$. Let $T_0^{q \times q}$ be such that

$$T_0 T_0' = (X_{n_0} X_{n_0}'); \quad \text{i.e., } T_0' (X_{n_0} X_{n_0}')^{-1} T_0 = I_q. \tag{6.6}$$

Writing

$$X_{N(0)}^{q \times (N - n_0))} = (\mathbf{x}_{n_0 + 1}, \ldots, \mathbf{x}_N), \tag{6.7}$$

we determine a matrix $A^{q \times (q + N - n_0)}$ (fixed according to some convention as soon as s_0^2 is fixed) such that

(i) $A(T_0, X_{N(0)}')' = I_q$

(ii) $AA' = zs_0^{-2}\Omega^{-1}$. (6.8)

By repeated application of Lemma 5.1, it can be shown that A subject to (6.8) can be found for N given by (6.5) (see Chatterjee 1962a). We then set up

$$\mathbf{l}_N = A(\hat{\beta}_0' T_0, y_{n_0+1}, \ldots, y_N)'. \tag{6.9}$$

Since $T_0'\hat{\beta}_0 \sim N_p(T_0'\beta, I_q)$ independently of s_0^2 and N becomes fixed as soon as s_0^2 is fixed, we get that given s_0^2, $l_n \sim N_q(\beta, zs_0^{-2}\Omega^{-1})$. From this, writing

$$Q = z^{-1}q^{-1}(\mathbf{l}_N - \beta)'\Omega(\mathbf{l}_N - \beta), \tag{6.10}$$

we get that $qs_0^2 Q$ is distributed as $\chi^2(q)$, given s_0^2, and hence, independently of s_0^2. We conclude the following.

THEOREM 6.1 *Q defined by (6.4)-(6.10) is distributed as $F(q, \nu_0)$.*

From this theorem it follows that if $z = d^2 q^{-1} F^{-1}(q, \nu_0; \alpha)$, then

$$(\beta - \mathbf{l}_N)'\Omega(\beta - \mathbf{l}_N) \leq d^2 \tag{6.11}$$

is a $(1-\alpha)$-confidence region for β. Again defining $Q^{(0)}$ by putting $\beta^{(0)}$ for β in (6.10), we can set up a right-tailed size α test on the basis of $Q^{(0)}$. The power of the test, which will be free from σ^2 and will involve the metric $z^{-1}(\beta-\beta^{(0)})'\Omega(\beta-\beta^{(0)})$, can be controlled by the choice of z.

As before, the above "exact" confidence region and test can be improved upon by replacing the "exactness" requirement by one of "boundedness." Thus starting with the same n_0, z, and Ω, after taking the first sample, we may define

$$N = \min\{n: n \geq n_0, (X_n X_n') - z^{-1}s_0^2\Omega \text{ is p.d.}\}. \tag{6.12}$$

Then writing

$$\hat{\beta}_N^* = (X_N X_N')^{-1}(X_N \tilde{Y}_N), \tag{6.13}$$

where N is as in (6.12) and $\tilde{Y}_n^{N\times 1} = (\tilde{y}_1, \ldots, \tilde{y}_N)'$, we can show that

$$(\beta - \hat{\beta}_N^*)'(X_N X_N')(\beta - \hat{\beta}_N^*) \leq qs_0^2 F(q, \nu_0; \alpha) \tag{6.14}$$

represents a $(1-\alpha)$-confidence region which lies completely within (6.11) when placed at the same center. Similarly, the test

$$\text{Reject } H_0 \text{ iff } q^{-1}s_0^{-2}(\hat{\beta}_N^* - \beta^{(0)})'(X_N X_N')(\hat{\beta}_N^* - \beta^{(0)}) > F(q, \nu_0; \alpha)$$

has size α and power uniformly greater than the size α test based on $Q^{(0)}$.

As before, the above procedures can be used to get point estimates of β with a stipulated expected loss, when the loss function is an increasing function of the metric $(\mathbf{a}-\beta)'\Omega(\mathbf{a}-\beta)$.

So far we have been considering the situation where we have an infinite sequence of observations $\tilde{y}_k$ corresponding to an infinite sequence of level combinations $\mathbf{x}_k$, $k = 1, 2, \ldots$, of q nonstochastic predictors $x_1, \ldots, x_q$. A more natural situation perhaps is where the predictors also are stochastic and $\tilde{y}$, $\tilde{x}_1, \ldots, \tilde{x}_q$ have a joint distribution for which the conditional distribution of $\tilde{y}$ given $x_1, \ldots, x_q$ is normal with a constant variance σ^2 and mean

$$E(\tilde{y} \mid x_1, \ldots, x_q) = \beta_1 x_1 + \cdots + \beta_q x_q. \tag{6.15}$$

(A constant term β_0 can be added to the right-hand side here with consequential changes in the procedure.) If we want to infer about $\beta_1, \ldots, \beta_q$ in (6.15), we can adopt any of the sampling schemes described above. However, since the observations on $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_q)'$ are not known beforehand, these will have to be made sequentially (at least as far as the predictors are concerned). Thus if one follows the scheme corresponding to (6.5), for each $n \geq n_0 + q$, after $\tilde{\mathbf{x}}_n$ is observed and $X_n X_n'$ computed, we have to check whether $X_n X_n' - z^{-1}s_0^2\Omega$ is p.d. If so, we stop; otherwise we proceed to observe $\tilde{\mathbf{x}}_{n+1}$. After the ultimate sample size N is determined and the $N - n_0$ additional units to be observed are selected in this way, $\tilde{y}_{n_0+1}, \ldots, \tilde{y}_N$ can all be observed at a time. (Of course, if observation of $\tilde{y}$ and $\tilde{\mathbf{x}}$ has to be synchronous, experimentation will have to be totally sequential.) In this case condition (6.3), ensuring the finiteness of N, has to be interpreted in an "almost sure" sense (for this it is sufficient if the dispersion matrix of $\tilde{\mathbf{x}}$ exists and is p.d.). While establishing the counterpart of Theorem 6.1, we have to conditionally fix s_0^2 as well as the infinite sequence $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots$, since N depends on both.

The version of the above problem with stochastic predictors is considered in Chatterjee (1962a). As noted, the case of nonstochastic predictors follows in an identical manner if the growth condition (6.3) on the predictor levels can be assumed. A confidence region analogous to (6.14) for nonstochastic predictors is also discussed inter alia in Mukhopadhyay and Abid (1986).

7. PROCEDURES OF STEIN'S TYPE FOR OTHER PROBLEMS AND SOME RELATED STUDIES

The idea of Stein's procedures was to carry out the sampling in two stages, the size of the second-stage sample being determined by the outcome of sampling in the first stage. This idea has been utilized to resolve difficulties created by the presence of unknown nuisance parameters in the context of certain other problems. Chapman (1950) suggested a two-stage solution to the Fisher–Behrens problem along similar lines: to compare the means of two normal populations with unknown and possibly unequal variances, one can use Stein's first procedure (Section 3) on each population to get two independent t-statistics whose difference would give an interval estimate and a valid test for the difference of the two means. A variant of this procedure based only on the t-distribution was proposed by B. K. Ghosh (1975). Again, for the problem of selection and ranking involving the means of several normal populations with a common unknown variance, a two-stage procedure was proposed by Bechhofer, Dunnett and Sobel (1954). In it, on the basis of an estimate of the common variance obtained from pilot samples, the number of

further observations to be taken is decided upon. The procedure is based on the multivariate t-distribution and is so designed that the probability of correct selection or ranking when the relevant differences in means exceed a given threshold value, is maintained above a specified lower bound. Multivariate analogues of two- and multi-population heteroscedastic problems have been tackled by Dudewicz and others, using the sampling scheme of Section 5.2 (see Dudewicz and Taneja 1987, where further references are given).

For problems involving estimation of several parameters (such as those in Sections 4–6), we discussed ellipsoidal confidence regions of bounded span and a predetermined confidence coefficient. By well-known techniques such confidence regions can be manipulated to yield simultaneous confidence intervals of bounded length for all normalized linear functions of the set of parameters (see Chatterjee 1962b and Healy 1956).

For point estimation of the mean vector of a multinormal population with all variances equal and all correlations 0, it is known that when the loss is given by the squared norm of errors and the number of variables is three or more, the sample mean vector can be improved by shrinking it towards a fixed point (James–Stein estimator; see e.g., Lehmann 1983, Section 4.5). Ghosh and Sen (1983) showed that a similar improvement can be effected in the overall sample mean if the sample is obtained in two stages and the loss is squared norm together with a "cost term" proportional to the total number of observations.

The argument underlying any two-stage procedure of Stein's type hinges on the fact of normality of the parent population at a crucial point. Normality ensures that even when the total number of observations is fixed by holding the first-stage estimate of dispersion constant, the distribution of the first sample mean, mean vector, or regression coefficients remains unaffected. It is natural to enquire whether departure from normality has any adverse effect on the performance of the procedure. Some robustness studies have been made for various types of departure from normality (see Ramkaran 1983, where further references are given) for the single mean problem. Generally, the finding is that the Stein confidence interval is remarkably robust against mild departures from normality both in terms of coverage probability and ASN.

Because of the crucial requirement that it must be possible to utilize the information about location given by the first sample, even when the estimate of the scale parameter from the first sample is conditionally fixed, the scope for developing exact two-stage procedures for nonnormal location-scale families seems to be limited. One notable result in this direction is that of Ghurye (1958), who considered the exponential distribution with density

$$f(y \mid \mu, \sigma) = 0 \qquad \text{for } y \le \mu,$$
$$= \sigma^{-1} e^{-\sigma^{-1}(y-\mu)} \qquad \text{for } y > \mu.$$

For this distribution Ghurye derived a two-stage test procedure for μ in which σ is estimated from a first sample $y_1, \ldots, y_{n_0}$ using $\sum_1^{n_0}(y_k - y_{(1)0})$, and μ is estimated from the overall sample of size N by $y_{(1)N}$, where $y_{(1)0}$ and $y_{(1)N}$ stand, respectively, for the smallest observation in the first sample and the overall sample.

8. A SHORTCOMING OF PROCEDURES OF STEIN'S TYPE: REMEDIES

As we emphasized in the introduction, Stein designed his sampling scheme, keeping comprehensive studentization in mind. As would be expected, in the process of achieving this objective the scheme does a certain amount of oversampling. Thus thinking in terms of Stein's second procedure for the interval estimation of a normal mean μ (Section 3.1), in the confidence interval given by (3.10)–(3.11), since an estimate of σ^2 based only on the first sample is used, a certain amount of information seems to be left unused. Alternatively, if we use the interval in the form $(\bar{\bar{y}}_N - d, \bar{\bar{y}}_N + d)$, the actual coverage probability would be well above the stated $1-\alpha$ level. All this suggests that some amount of sampling could be avoided had we utilized all the information available. We can see this concretely by comparing the ASN of the procedure with the fixed sample size $n = n^*(\sigma^2)$ required to achieve the same objective when σ^2 is known. If $\pm\tau(\alpha/2)$ are the two-sided α-points of the standard normal distribution, then, knowing σ^2, we require $n^{-1/2}\tau(\alpha/2)\sigma = d$, to get a confidence interval of span $2d$, which means

$$n = n^*(\sigma^2) = d^{-2}\tau^2(\alpha/2)\sigma^2. \tag{8.1}$$

On the other hand, from (3.6) and (3.25) for the Stein procedure, if n_0 and σ^2/d^2 are both small,

$$E_{\sigma^2}(N) \simeq z^{-1}\sigma^2 = d^{-2}t^2(\nu_0; \alpha/2)\sigma^2.$$

Thus

$$\frac{E_{\sigma^2}(N)}{n^*(\sigma^2)} \simeq \frac{t^2(\nu_0; \alpha/2)}{\tau^2(\alpha/2)}. \tag{8.2}$$

In fact, using (3.22)–(3.24) we can show that

$$\lim_{d\to 0} \frac{E_{\sigma^2}(N)}{n^*(\sigma^2)} = \frac{t^2(\nu_0; \alpha/2)}{\tau^2(\alpha/2)} < 1. \tag{8.3}$$

This fact is usually expressed by saying that the Stein procedures are *asymptotically inefficient* (see Chow and Robbins 1965 and Ghosh and Mukhopadhyay 1981).

To get round this difficulty, one could take to full-scale sequential sampling as proposed by Chow and Robbins (1965) in the case of a general population with finite variance (see Chapters 10 and 11). Thereby one can get a procedure for which $\mathrm{ASN}/n^*(\sigma^2) \to 1$ as $d \to 0$. However, that would involve one-at-a-time sampling. Besides, the coverage probability of the resulting interval would be $1-\alpha$ only in the limit, so for a given $d > 0$ one can at best have a vague assurance that the same is close to the stated level.

Is it possible to remedy this shortcoming within the framework of the Stein procedure itself? A close examination of (3.8) shows that the solution lies in the proper choice of the pilot sample size n_0. Thus if n_0 is about $z^{-1}\sigma^2$, N in (3.8) will

be around n_0 and not much information will be left out if σ^2 is estimated solely from the pilot sample. In fact if one gives a formula for the choice of n_0 such that as $d \to 0$, $n_0 \to \infty$ at an appropriate rate, one can achieve "asymptotic efficiency" in the sense that

$$\lim_{d\to\infty} \frac{E_{\sigma^2}(N)}{n^*(\sigma^2)} = 1. \tag{8.4}$$

This forms the basis of the modified two-stage procedures as considered in Ghosh and Mukhopadhyay (1981) and Mukhopadhyay (1980).

The crux of the problem, however, is not the theoretical satisfaction of realizing (8.4). Even if (8.4) holds in the case of a rule for choosing n_0, for a particular $d > 0$, σ^2 may be so large as to make $E_{\sigma^2}(N)/n^*(\sigma^2)$ substantially larger than 1. One way of resolving this difficulty would be to take an initial (prepilot) sample of small size to get a rough idea about σ^2 and to decide upon the size of the pilot sample on that basis. Thus we may take a prepilot sample $\tilde{y}_1, \ldots, \tilde{y}_{m_0}$ of size m_0 and estimate σ^2 by $s^2_{(0)}$ as usual on its basis. Let

$$N_0 = N_0(s^2_{(0)}) = \max\left\{m_0 + 1, \left[cd^{-2}\tau^2(\alpha/2)s^2_{(0)}\right] + 1\right\} \tag{8.5}$$

where $0 < c < 1$ is some fixed number. We take $N_0 - m_0$ additional observations $\tilde{y}_{m_0+1}, \ldots, \tilde{y}_{N_0}$ and denoting

$$\tilde{\bar{y}}_{(0)} = m_0^{-1}\sum_1^{m_0} \tilde{y}_k, \quad \tilde{\bar{y}}_{(1)} = \nu_{(1)}^{-1}\sum_{m_0+1}^{N_0} \mathrm{y}_k, \qquad \nu_{(1)} = N_0 - m_0$$

we find

$$s^2_{(1)} = \nu_{(1)}^{-1}\left\{\sum_{m_0+1}^{N_0} (\tilde{y}_k - \tilde{\bar{y}}_{(1)})^2 + m_0\nu_{(1)}N_0^{-1}(\tilde{\bar{y}}_{(0)} - \tilde{\bar{y}}_{(1)})^2\right\}. \tag{8.6}$$

Note that, given $s^2_{(0)}$, N_0 and $\nu_{(1)}$ are fixed and $\nu_{(1)}s^2_{(1)}/\sigma^2$ is conditionally distributed as a $\chi^2(\nu_{(1)})$ independently of $\sum_1^{N_0}\tilde{y}_k$. We next determine

$$N = \max\left\{N_0, \left[d^{-2}t^2_{(1)}s^2_{(1)}\right] + 1\right\}, \tag{8.7}$$

where we write $t_{(1)}$ for $t(\nu_{(1)}; \alpha/2)$. We then take, if necessary, $N - N_0$ additional observations $\tilde{y}_k$, $k = N_0 + 1, \ldots, N$, and find the overall mean $\tilde{\bar{y}}_N = N^{-1}\sum_1^N \tilde{y}_k$. Arguing as in Section 3.1, we can show that conditionally for fixed $s^2_{(0)}$, and hence unconditionally, $(\tilde{\bar{y}}_N - d, \tilde{\bar{y}} + d)$ covers μ with probability strictly greater than $1-\alpha$. The idea behind this proposal is to make the pilot sample size N_0 commensurate with $s^2_{(0)}$, which is an unbiased estimate of σ^2. Note that m_0 being a fixed number, by (8.5), for small d, $N_0 \simeq cd^{-2}\tau^2(\alpha/2)s^2_{(0)}$. Using this and (8.1) and arguing as in

(3.24), we get

$$\frac{E_{\sigma^2}(N \mid s^2_{(0)})}{n^*(\sigma^2)} \simeq c\frac{s^2_{(0)}}{\sigma^2} P\left\{\chi^2(\nu_{(1)}) \le \nu_{(1)}c\tau^2(\tfrac{1}{2}\alpha)t^{-2}_{(1)}s^2_{(0)}\sigma^{-2}\right\} + \tau^{-2}(\alpha/2)t^2_{(1)}P\left\{\chi^2(\nu_{(1)}+2) > \nu_{(1)}c\tau^2(\alpha/2)t^{-2}_{(1)}s^2_{(0)}\sigma^{-2}\right\}.$$

For fixed $s^2_{(0)}$ as $d \to 0$, $\nu_{(1)} \to \infty$ and $t_{(1)} \to \tau(\alpha/2)$. Hence using the fact that as $\nu \to \infty$, $(2\nu)^{-1/2}\{\chi^2(\nu) - \nu\}$ is asymptotically distributed as $N(0,1)$, we can show that

$$E_{\sigma^2}\frac{(N \mid s^2_{(0)})}{n^*(\sigma^2)} \simeq \frac{cs^2_{(0)}}{\sigma^2} \qquad \text{if } cs^2_{(0)} > \sigma^2$$
$$\simeq 1 \qquad \text{if } cs^2_{(0)} \le \sigma^2.$$

Since $P\left\{cs^2_{(0)} > \sigma^2\right\} = P\left\{\chi^2(m_0 - 1) > (m_0 - 1)c^{-1}\right\}$ would be small if c is well below 1 (say, $c \simeq 1/2$), the efficiency of the above procedure would be close to 1.

Although the above three-stage procedure gives a confidence interval of span $2d$ and confidence coefficient at least $1 - \alpha$ with asymptotic efficiency close to 1, it is admittedly somewhat crude: in estimating σ^2 for the determination of N in (8.7) we ignore the information in the prepilot sample. Hall (1981) proposed a more refined three-stage procedure in which, after determining $s^2_{(0)}$ from a prepilot sample of size m_0 as above, we find

$$N_1 = \max\left\{m_0, \left[cd^{-2}\tau^2(\alpha/2)s^2_{(0)}\right] + 1\right\}, \tag{8.8}$$

$0 < c < 1$ being as before a fixed number. Then $N_1 - m_0$ additional observations are taken if $N_1 > m_0$, and the usual estimate $s^2_{N_1}$ of σ^2 is computed from all the N_1 observations. The ultimate sample size is

$$N = \max\left\{N_1, \left[d^{-2}\tau^2(\alpha/2)s^2_{N_1} + (1/2)c^{-1}\left\{5 + \tau^2(\alpha/2) - c\right\}\right] + 1\right\} \tag{8.9}$$

An approximate $(1-\alpha)$-confidence interval is given by $(\tilde{\bar{y}}_N - d, \tilde{\bar{y}}_N + d)$. Hall has shown that the actual coverage probability of this interval is strictly greater than $1 - \alpha$ for d sufficiently small, and, further, for N given by (8.8),

$$E_{\sigma^2}(N) = n^*(\sigma^2) + (1/2)c^{-1}\left\{\tau^2(\alpha/2) + 1\right\} + o(1).$$

Obviously, as $d \to 0$, $E_{\sigma^2}(N)/n^*(\sigma^2) \to 1$. This procedure is only marginally more costly than the sequential procedure of Chow and Robbins.

Versions of Hall's three-stage sampling procedure for other problems are discussed in Mukhopadhyay and Abid (1986) and Mukhopadhyay and Al-Mousawi (1986).

REFERENCES

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.

Chapman, D. G. (1950). Some two-sample tests. *Ann. Math. Statist. 21*, 601–606.

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

Chatterjee, S. K. (1959b). Some further results on the multinormal extension of Stein's two-sample procedure. *Calcutta Statist. Assoc. Bull. 9*, 20–28.

Chatterjee, S. K. (1960). Some sequential (two-step and multi-step) inference procedures with "nuisance-parameter-free" performance in multivariate problems. Ph.D. thesis (unpublished), accepted by Calcutta University.

Chatterjee, S. K. (1962a). Sequential inference procedures of Stein's type for a class of multivariate regression problems. *Ann. Math. Statist. 33*, 1039–1064.

Chatterjee, S. K. (1962b). Simultaneous confidence intervals of predetermined length based on sequential samples. *Calcutta Statist. Assoc. Bull. 11*, 144–159.

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.

Dantzig, G. B. (1940). On the nonexistence of test of "Student's" hypothesis having power functions independent of σ. *Ann. Math. Statist. 11*, 186–192.

Dudewicz, E. J. and Taneja, V. S. (1987). The heteroscedastic method: multivariate implementation. *Ann. Inst. Statist. Math. 39*, 177–190.

Ghosh, B. K. (1975). A two-stage procedure for the Behrens–Fisher problem. *J. Amer. Statist. Assoc. 70*, 457–462.

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

Ghosh, M. and Sen, P. K. (1983). On two-stage James–Stein estimators. *J. Sequential Anal. 2*, 359–367.

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

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

Healy, W. C., Jr. (1956). Two-sample procedures in simultaneous estimation. *Ann. Math. Statist. 27*, 687–702.

Hewett, J. D. and Spurrier, J. D. (1983). A survey of two-stage tests of hypotheses: theory and application. Commun. Statist.—*Theor. Meth. 12* (20), 2307–2425.

Lehmann, E. L. (1950). Theory of Estimation. Notes recorded by C. Blyth, Associated Student Store, University of California.

Lehmann, E. L. (1983). *Theory of Point Estimation*, Wiley, New York.

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. and Abid, A. D. (1986). On fixed size confidence regions for the regression parameters. *Metron. 34*, 297–306.

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

Ramkaran. (1983). The robustness of Stein's two-stage procedure. *Ann. Statist. 11*, 1251–1256.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications.* Wiley Eastern, New Delhi.

Ruben, H. (1961). Studentization of two-stage sample means from normal populations with unknown common variance. *Sankhyā. 23*, 231–250.

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

Wald, A. (1947). *Sequential Analysis.* Wiley, New York.