NOTE ON SUFFICIENT STATISTICS AND TWO-STAGE PROCEDURES

By S. G. Ghurye

University of Chicago

0. Introduction. This note is the result of an attempt to discover problems in which one can apply the two-stage procedure used by Stein [1] for tests regarding the mean of a normal population. One such problem, that of testing for a location parameter of an exponential population, was found to be easily soluble along the lines of Stein's work. An investigation of the problem of optimum statistics for such procedures was also undertaken, and partial solutions, given in Sec. 2, were found. In this connection, the author would like to thank the referee for his useful comments.

1. Testing for a location parameter of a distribution. Throughout this paper $F(x)$ will be a one-dimensional c.d.f. with at least two points of increase. Further, $\{Y_n\}$ will always denote a sequence of independent random variables having a common c.d.f. $F(x)$ and $\{X_n\}$ will denote a family of sequences of independent random variables, all elements of any one sequence having a common c.d.f. $F[(x - \theta)/\sigma]$, $-\infty < \theta < \infty$, $\sigma > 0$. We shall be dealing with statistics or sequences of real and single-valued functions $t(n; x_1, \cdots, x_n)$ and $s(n; x_1, \cdots, x_n)$ of n real variablès, $n = 1, 2, \cdots$, about which one or more of the following assumptions will be made as required:

Assumption I. For any integer $n > 0$, any $a > 0$, any real b and any

$$(x_1, \cdots, x_n) \, \varepsilon \, R^n,$$

$$t(n; ax_1 + b, \cdots, ax_n + b) = at(n; x_1, \cdots, x_n) + b. \tag{1}$$

Assumption II. Analogously,

$$s(n; ax_1 + b, \cdots, ax_n + b) = as(n; x_1, \cdots, x_n). \tag{2}$$

Assumption III. There exists a positive, nondecreasing and unbounded sequence $k(n)$ such that

$$\Pr \{t(n; Y_1, \cdots, Y_n) \leqq x/k(n)\} = G(x) \tag{3}$$

is independent of n. Without loss of generality, we may assume $k(1) = 1$.

Assumption IV. The random variables $t(n; Y_1, \cdots, Y_n)$ and $s(n; Y_1, \cdots, Y)_n$ are stochastically independent.

Assumption V. There exists a positive integer m, such that for any $n > m$, $t(n; x_1, \cdots, x_n)$ is a function only of m, n, $t(m; x_1, \cdots, x_m)$ and $x_{m+1}, \cdots, x_n$.

Received January 16, 1956; revised March 18, 1957.

ASSUMPTION VI. Let

$$\Pr\{s(n; Y_1, \cdots, Y_n) \leqq x\} = H(x; n). \tag{4}$$

Then $H(0; n) = 0$ for all n.

Now, let $\mathcal{P}$ be a population whose c.d.f. is known to be $F[(x - \theta)/\sigma]$, but θ, σ are unknown, and suppose that it is desired to obtain a test of $H_0: \theta = \theta_0$ against the alternative $\theta > \theta_0$ with the following properties:

(a) The size of the test is to be a prescribed probability α for all $\sigma > 0$;

(b) The power of the test for $\theta = \theta_0 + \delta$, where $\delta > 0$ is a given number, must be not less than a prescribed probability $\beta > \alpha$ for all $\sigma > 0$;

(c) The power $\rightarrow 1$ as $\theta \rightarrow \infty$.

It is easy to see that, if Assumptions I–VI are satisfied, the following procedure, which is essentially that given by Stein, has the required properties:

Choose an m satisfying Assumption V, and let

$$\chi = \text{infimum of all } y \text{ such that } \int_{-\infty}^{\infty} G(yu)\, dH(u; m) \geqq 1 - \alpha; \tag{5}$$

$$\gamma = \begin{cases} \left[\int_{-\infty}^{\infty} G(\chi u)\, dH(u; m) - 1 + \alpha\right] \Big/ \int_{-\infty}^{\infty} \{G(\chi u) - G(\chi u - 0)\}\, dH(u; m) \\ \qquad\qquad\qquad\qquad \text{if the denominator} > 0, \\ 0 \qquad \text{otherwise;} \end{cases} \tag{6}$$

$$P(y) = 1 - \int_{-\infty}^{\infty} G(yu)\, dH(u; m) + \gamma \int_{-\infty}^{\infty} \{G(yu) - G(yu - 0)\}\, dH(u; m); \tag{7}$$

$$\chi' = \text{supremum of all } y \text{ such that } P(y) \geqq \beta; \tag{8}$$

$$\rho = (\chi - \chi')/\delta > 0. \tag{9}$$

Take m independent observations $X_1, \cdots, X_m$ from $\mathcal{P}$, and calculate

$$s_m = s(m; X_1, \cdots, X_m).$$

Let N be such that

$$k(N - 1) < \rho s_m \leqq k(N), \tag{10}$$

except if $\rho s_m < k(m)$, in which case $N = m$.

If $N > m$, take $N - m$ more independent observations $X_{m+1}, \cdots, X_N$ from $\mathcal{P}$, calculate

$$U = \{t(N; X_1, \cdots, X_N) - \theta_0\} k(N)/s_m, \tag{11}$$

and reject H_0 with probability $\phi(U)$, where

$$\phi(u) = \begin{cases} 0, & u < \chi, \\ \gamma, & u = \chi, \\ 1, & u > \chi. \end{cases} \tag{12}$$

The expected sample-size is

$$E(N) = m \Pr \{\rho s_m \leqq k(m)\} + \sum_{r=1}^{\infty} (m + r) \Pr \{k(m + r - 1) < \rho s_m \leqq k(m + r)\} \tag{13}$$

$$= m + \sum_{r=m}^{\infty} \bar{H}\{k(r)\sigma^{-1}\rho^{-1}; m\},$$

where $\bar{H}(x) = 1 - H(x)$.

For computational convenience, we have the inequalities

$$\nu < E(N) < \nu + \epsilon, \tag{14}$$

where

$$\nu = mH\{k(m)\sigma^{-1}\rho^{-1}; m\} + \int_{\sigma\rho u > k(m)} k^{-1}(\sigma\rho u)\, dH(u; m), \tag{15}$$

$$\epsilon = \bar{H}\{k(m)\sigma^{-1}\rho^{-1}; m\}, \tag{16}$$

and $k^{-1}(u)$ is any monotone function of $u > 0$ such that $k\{k^{-1}(n)\} = n$ for every integer $n > 0$.

It may be noted that, if in Assumption III we drop the restriction $k(1) = 1$, then $ck(n)$, with $c > 0$, serves instead of $k(n)$ (with a different G for each c). However from (10) it is easy to see that N is independent of this c. Thus the restriction $k(1) = 1$ does not cause any loss of generality. In the same way, it can be seen that if we substitute $cs(n; x_1, \cdots, x_n)$ for $s(n; x_1, \cdots, x_n)$, $c > 0$, N is unaffected.

Examples in which t and s satisfying the assumptions can be found are provided by the normal distribution, which was discussed in detail by Stein, and the exponential distribution which we shall take up here.

Of the several possible choices for (t, s) in the normal case, Stein considered two, in both of which s^2 is the usual estimate of σ^2. By using a special linear function for t, he was able to obtain a test whose power is independent of σ instead of merely being bounded below by a function independent of σ as required in property (b) above. However, he noted that this procedure "wastes information," and advocated one using the sample mean as t. In fact, the use of any statistic other than the sample mean is wasteful in the sense that it leads to a higher expected sample-size, as we shall see in Sec. 2.

Next, let $\mathcal{P}$ be a population with c.d.f. $F[(x - \theta)/\sigma]$, where

$$F(x) = \begin{cases} 0 & \text{if } x \leqq 0, \\ 1 - e^{-x} & \text{if } x > 0. \end{cases} \tag{17}$$

Let $x_{[1]}, \cdots, x_{[n]}$ denote the rearrangement of numbers $x_1, \cdots, x_n$ in ascending order of magnitudes, and let

$$t_0(n; x_1, \cdots, x_n) = x_{[1]} = \min (x_1, \cdots, x_n). \tag{18}$$

This is a sufficient estimator of θ, and we shall see in the next section that it is the best for use as "t". If corresponding to independent observations $X_1, \cdots, X_n$ from $\mathcal{P}$, we put $Z_1 = X_{[1]}$, and

$$Z_i = X_{[i]} - X_{[i-1]}, \qquad i = 2, 3, \cdots, n,$$

the joint density of $Z_1, \cdots, Z_n$ is

$$(19)\quad f(z_1, \cdots, z_n) = \begin{cases} n!\, \sigma^{-n} \exp\{-n(z_1 - \theta) - (n-1)z_2 - \cdots - z_n\}/\sigma \\ \qquad\qquad \text{if } z_1 \geqq \theta, z_i \geqq 0, i = 2, 3, \cdots n, \\ 0 \quad \text{otherwise.} \end{cases}$$

Any function satisfying Assumption II is a function only of the differences of the arguments, and hence only of $Z_2, \cdots, Z_n$. It is thus independent of Z_1 and can be used as "s" if it is positive with probability 1. This is true more generally, as shown by Lemma 2 of the next section. It will be seen (Example 2) that, asymptotically as $\sigma \to \infty$, the best statistic to use as s is

$$(20)\qquad s_0(n; x_1, \cdots, x_n) = \sum_1^n x_i - nt_0(n; x_1, \cdots, x_n),$$

which together with t_0 is sufficient for σ.

For this pair of statistics, we have from (19)

$$(21)\qquad k(n) = n, \qquad G(x) = F(x),$$

$$(22)\qquad H(u; m) = \begin{cases} \int_0^u x^{m-2} e^{-x}\, dx/(m-2)!, & u > 0 \\ 0, & u \leqq 0; \end{cases}$$

$$(23)\qquad \begin{aligned} \gamma &= 0, \qquad \chi = \alpha^{-1/(m-1)} - 1, \\ \delta\rho &= \{\alpha^{-1/(m-1)} - \beta^{-1/(m-1)}\}, \end{aligned}$$

and ν, ϵ are given by

$$(24)\qquad \begin{aligned} \nu &= m \int_0^{mc} u^{m-2} e^{-u}\, du/(m-2)! + \{(m-1)/c\} \int_{mc}^{\infty} u^{m-1} e^{-u}\, du/(m-1)! \\ \epsilon &= \int_{mc}^{\infty} u^{m-2} e^{-u}\, du/(m-2)!, \end{aligned}$$

where $c = (\sigma\rho)^{-1}$.

The values of ν and ϵ were calculated for $\alpha = 0.05 = 1 - \beta$ and $\alpha = 0.01 = 1 - \beta$ and several values of m and δ/σ. These are given in Table 1.

2. Optimum choice of statistics. We shall now prove three preliminary lemmas which enable us to show that if a suitable sufficient estimator of θ exists, it minimizes the expected sample size among all t satisfying the assumptions.

LEMMA 1. *Let Y be a real-valued, one-dimensional random variable, and $f(y)$ a*

TABLE 1

Expected sample size

(The entries are given in the form $\nu + \epsilon$ and imply that $\nu < E(N) < \nu + \epsilon$)

		δ/σ						
	m	0.05	0.10	0.20	0.40	0.60	0.80	1.00
	2	379.0	189.4	94.8	47.4	31.6	23.8	19.1
		+1.0	+1.0	+1.0	+1.0	+0.9	+0.9	+0.9
	4	101.8	50.9	25.5	12.8	8.7	6.4	5.7
		+1.0	+1.0	+1.0	+0.9	+0.8	+0.7	+0.6
	6	81.0	40.5	20.3	10.4	7.6	6.5	6.2
$\alpha = .05$		+1.0	+1.0	+1.0	+0.8	+0.5	+0.3	+0.1
	8	73.6	36.9	18.5	10.0	8.3	8.1	8.0
		+1.0	+1.0	+1.0	+0.7	+0.2	+0.0	+0.0
	10	70.0	35.0	17.6	10.7	10.0	10.0	10.0
		+1.0	+1.0	+0.9	+0.3	+0.0	+0.0	+0.0
	20	63.9	32.3	20.3	20.0	20.0	20.0	20.0
		+1.0	+1.0	+0.1	+0.0	+0.0	+0.0	+0.0
	2	1980.0	990.0	495.0	247.5	165.5	123.8	99.0
		+1.0	+1.0	+1.0	+1.0	+1.0	+1.0	+1.0
	4	218.3	109.1	54.6	27.3	18.2	13.7	11.0
		+1.0	+1.0	+1.0	+1.0	+1.0	+0.9	+0.9
	6	151.0	75.5	37.8	18.9	12.6	9.8	8.2
$\alpha = .01$		+1.0	+1.0	+1.0	+0.9	+0.9	+0.8	+0.7
	8	130.1	65.0	32.6	16.3	11.3	9.3	8.2
		+1.0	+1.0	+1.0	+0.9	+0.7	+0.5	+0.2
	10	120.6	60.3	30.0	15.3	11.3	10.3	10.0
		+1.0	+1.0	+1.0	+0.8	+0.5	+0.2	+0.0
	20	104.0	52.0	26.8	20.0	20.0	20.0	20.0
		+1.0	+1.0	+0.8	+0.0	+0.0	+0.0	+0.0

measurable, real-valued function with the property that for any real x, θ and any $\sigma > 0$,

$$\Pr\{f(\sigma Y + \theta) - \theta \leqq \sigma x\} = \Pr\{f(Y) \leqq x\}. \tag{25}$$

Then if $f(y)$ is strictly monotone, there exists an interval I, open or closed, such that $y \varepsilon I \Rightarrow f(y) = y$, and $\Pr\{Y \varepsilon I\} = 1$.

Proof. To start with, we note that since the right hand member of (25) is a nondecreasing function of x, $f(y)$ cannot be a decreasing function; for if it were, we would have

$$\Pr\{f(Y) \leqq x\} = \Pr\{f(Y + \theta) - \theta \leqq x\} \geqq \Pr\{f(Y) - \theta \leqq x\}$$

for all $\theta > 0$.

This implies that the c.d.f. of Y is constant and hence contradicts the assumption that $F(x)$ has at least two points of increase.

Therefore $f(y)$ is increasing, and

$$\{y:f(y) \leqq u\} = \{y:y \leqq f^{-1}(u)\}.$$

Let $h(y) = f^{-1}(y) - y$; then from (25) we obtain

$$\Pr\left\{Y < x + \frac{h(\sigma x + \theta)}{\sigma}\right\} = \Pr\{Y < x + h(x)\}, \tag{26}$$

$$\Pr\left\{Y = x + \frac{h(\sigma x + \theta)}{\sigma}\right\} = \Pr\{Y = x + h(x)\}. \tag{27}$$

Suppose there exists a y_0 such that $h(y_0) \neq 0$. From (26), we get the relation

$$\Pr\left\{Y < y_0 + \frac{h(y_0)}{\sigma}\right\} = \Pr\{Y < y_0 + h(y_0)\}.$$

By letting $\sigma \to 0$ and again $\sigma \to \infty$, we see that

$$\begin{cases} h(y_0) > 0 \quad \text{implies} \quad \Pr\{Y \leqq y_0\} = 1, \\ h(y_0) < 0 \quad \text{implies} \quad \Pr\{Y < y_0\} = 0. \end{cases} \tag{28}$$

Hence, if there exist x_0, y_0 such that $h(x_0) < 0$ and $h(y_0) > 0$, we must have $x_0 < y_0$. Consequently, there exist points x_0, y_0 (which may be respectively $-\infty$ and ∞) such that

$$\begin{cases} h(x_0) \leqq 0, \quad h(y_0) \geqq 0, \quad h(y) = 0 \quad \text{for} \quad x_0 < y < y_0, \\ \qquad\qquad \text{and} \quad \Pr\{x_0 \leqq Y \leqq y_0\} = 1. \end{cases} \tag{29}$$

Finally, if $h(y_0) > 0$ we see from (27), by choosing $\sigma = 1$ and θ such that $x_0 - y_0 < \theta < 0$, that

$$\Pr\{Y = y_0\} = \Pr\{Y = y_0 + h(y_0)\} = 0$$

from (29), and similarly, $h(x_0) < 0$ implies $\Pr\{X = x_0\} = 0$. Hence the result.

Next we want to consider two statistics, one of which is sufficient for θ and both of which have θ as a location parameter. More specifically we prove

Lemma 2. *Let $P(\cdot\,;\theta)$, $-\infty < \theta < \infty$, be a family of probability measures on a countably additive class of subsets of a set Ω of points ω; let $f(\omega)$, $g(\omega)$ be measurable real valued functions on Ω such that for any Borel sets S, T on the real line,*

$$P\{f^{-1}(S + \theta) \cap g^{-1}(T + \theta); \theta\} \underset{\theta}{\equiv} P\{f^{-1}(S) \cap g^{-1}(T); 0\}. \tag{30}$$

If $f(\omega)$ is a sufficient statistic for the family $P(\cdot\,;\theta)$, the random variables $g(\omega) - f(\omega)$ and $f(\omega)$ are stochastically independent.

Proof. Writing $Pf^{-1}(S)$ to denote $P\{f^{-1}(S) \cap \Omega; 0\}$, we have from (30)

$$P\{f^{-1}(S) \cap \Omega; \theta\} = Pf^{-1}(S - \theta). \tag{31}$$

By the Radon-Nikodym Theorem and the sufficiency of $f(\omega)$, we know that corresponding to each set T, there exists an integrable function $\lambda(T \mid x)$ on the real line such that

$$P\{f^{-1}(S) \cap g^{-1}(T); \theta\} \underset{\theta}{\equiv} \int_S \lambda(T \mid x)\, dPf^{-1}(x - \theta).$$

This gives us

$$\begin{aligned} P\{f^{-1}(S) \cap g^{-1}(T); 0\} &= P\{f^{-1}(S + \theta) \cap g^{-1}(T + \theta); \theta\} \\ &= \int_S \lambda(T + \theta \mid x + \theta)\, dPf^{-1}(x). \end{aligned}$$

It follows that for every set T, we have

$$\begin{aligned} P\{f^{-1}(S) \cap g^{-1}(T); 0\} &= \int_S \lambda(T - x \mid 0)\, dPf^{-1}(x) \\ &= \int_S \mu(T - x)\, dPf^{-1}(x), \end{aligned}$$

so that

$$\Pr\{g(\omega)\ \varepsilon\ T \mid f(\omega) = x\} = \mu(T - x) \text{ for a.e. } x\ [Pf^{-1}],$$

and consequently

$$\Pr\{g(\omega) - f(\omega)\ \varepsilon\ T \mid f(\omega) = x\} = \mu(T)$$

is independent of x.

COROLLARY 2.1. *Let* $t_j(n; x_1, \cdots, x_n), j = 0, 1,$ *be functions satisfying Assumption* I *and let* $t_0(n; x_1, \cdots, x_n)$ *be a sufficient statistic for the family of distributions* $\prod_{j=1}^{n} F(x_j - \theta), -\infty < \theta < \infty$. *Then for any n, if* $X_1, \cdots, X_n$ *are independent random variables having the common* c.d.f $F[(x - \theta) / \sigma]$, *the random variables* $t_0(n; X_1, \cdots, X_n)$ *and* $t_1(n; X_1, \cdots, X_n) - t_0(n; X_1, \cdots, X_n)$ *are independent.*

COROLLARY 2.2. *If* $t_0(n; x_1, \cdots, x_n)$ *is as in Corollary* 2.1 *and* $s(n; x_1, \cdots, x_n)$ *is any function satisfying Assumption* II, *the random variables* $t_0(n; X_1, \cdots, X_n)$ *and* $s(n; X_1, \cdots, X_n)$ *are independent.*

LEMMA 3[1]. *Let* $t_0, t_1, X_1, \cdots, X_n$ *be as in Corollary* 2.1 *and suppose that* t_0, t_1 *also satisfy Assumption* III *with respective sequences* $k_0(n), k_1(n)$. *Then*

$$k_0(n) \geqq k_1(n), \tag{32}$$

the equality holding if and only if

$$\Pr\{t_0(n; X_1, \cdots, X_n) = t_1(n; X_1, \cdots, X_n)\} = 1. \tag{33}$$

Further, for any a, b such that $a < 0 < b$,

$$\begin{aligned} &\Pr\{a < t_0(n; X_1, \cdots, X_n) - \theta < b\} \\ &\qquad \geqq \Pr\{a < t_1(n; X_1, \cdots, X_n) - \theta < b\}. \end{aligned} \tag{34}$$

[1] It may be of interest to compare this with the results of Pitman [3], pp. 401–402.

PROOF. From Assumptions I and III, it follows that

$$\Pr\{t_j(n; X_1, \cdots, X_n) \leqq x\} = F\left\{\frac{(x-\theta)}{\sigma} k_j(n)\right\}, \qquad j = 0, 1. \tag{35}$$

Let $f(z) = \int e^{izx}\, dF(x), \qquad c_n = k_0(n)/k_1(n)$, and

$$g\{z/k_0(n)\} = E\{e^{iz[t_1(n;\, Y_1, \cdots, Y_n) - t_0(n;\, Y_1, \cdots, Y_n)]}\}.$$

Then from Corollary 2.1 and (35), we get

$$f(c_n z) = f(z)g(z). \tag{36}$$

First suppose $f(z_0) = 0$ for some z_0. Then $f(c_n^r z_0) = 0$ for every r, and hence $c_n \nless 1$. On the other hand, if $f(z) \neq 0$ for all z, $g(z) = f(c_n z)/f(z)$ is a characteristic function. Hence $f(c_n^r z)/f(z) = g(z)g(c_n z) \cdots g(c_n^{r-1} z)$, $r = 1, 2, \cdots$, is a sequence of characteristic functions. If $c_n < 1$, $\lim_{r\to\infty} f(c_n^r z)/f(z) = 1/f(z)$ for every finite z, and since the limit is continuous in z, it is a characteristic function. But $f(z)$ is also a characteristic function. This implies $|f(z)| = 1$ for all z, which is impossible if F has more than one point of increase. Consequently, $c_n \geqq 1$, that is to say (32) holds.

It follows from (36) that $c_n = 1$ if and only if $g(z) = 1$, in which case (33) follows on account of Assumption I.

Finally, (34) is an immediate consequence of (32) and (35).

THEOREM 1. *Let t_1 and s be statistics satisfying Assumptions* I–VI, *and let t_0 be a statistic, satisfying Assumptions* I, III *and* V, *which is sufficient for the family of distributions $\Pi_1^n F(x_j - \theta)$, $-\infty < \theta < \infty$. Let N_i denote the sample-size in the two-stage procedure using (t_i, s), $i = 0, 1$, with the same m, α, β, and δ. Then*

$$E(N_0) \leqq E(N_1) \quad \text{for} \quad \text{all} \quad \sigma, \tag{37}$$

the equality holding for all σ if and only if

$$\Pr\{t_0(n; X_1, \cdots, X_n) = t_1(n; X_1, \cdots, X_n)\} = 1 \text{ for all } n \geqq m.$$

PROOF. The hypotheses of the theorem and Corollary 2.2 enable us to use (t_0, s) for the procedure described in the previous section. With the notation used there we have from (30), $G_i(x) = F(x)$, $i = 0, 1$, and from (9),

$$\rho_0 = \rho_1 . \tag{38}$$

From (13), we get

$$E(N_i) = m + \sum_{r=m}^{\infty} \bar{H}\{k_i(r)\sigma^{-1}\rho_i^{-1}; m\},$$

and using (32) and (38) the result follows.

REMARK. Theorem 1 solves part of the problem of optimization of the two-sample procedure by showing that if a suitable sufficient estimator of θ exists, it is the best "t" to use. This leaves us with the problem of choosing "s". We shall see that in the case of the normal and exponential distributions the best pair (t, s) to use, asymptotically as $\sigma \to \infty$, is the pair of sufficient statistics (t_0, s_0).

LEMMA 4[2] *Let* $s_0(n; x_1, \cdots, x_n)$ *and* $s(n; x_1, \cdots, x_n)$ *be statistics satisfying Assumptions* II *and* VI, *and let* t_0 *be a statistic satisfying Assumption* I. *Let* (t_0, s_0) *be sufficient for the family* $\Pi_1^n F[(x_i - \theta)/\sigma]$, $-\infty < \theta < \infty$, $\sigma > 0$. *Then* $t_0(n; X_1, \cdots, X_n)$, $s_0(n; X_1, \cdots, X_n)$, *and* $s(n; X_1, \cdots, X_n)/s_0(n; X_1, \cdots, X_n)$ *are mutually independent.*

PROOF. This result can be proved formally along the lines used for Lemma 2, but this seems hardly necessary, and only an outline in terms of conditional probabilities will be given.

Let $u(n; x_1, \cdots, x_n) = s(n; x_1, \cdots, x_n)/s_0(n; x_1, \cdots, x_n)$, and note that u is invariant under the transformation $x_i \rightarrow \sigma x_i + \theta$, $i = 1, \cdots, n$.

For almost all a and b,

$$\Pr\{u(n; X_1, \cdots, X_n) \in S \mid t_0(n; X_1, \cdots, X_n) = a, \quad s_0(n; X_1, \cdots, X_n) = b\} \tag{39}$$

is independent of (θ, σ) and equals

$$\Pr\{u(n; Y_1, \cdots, Y_n) \in S \mid t_0(n; Y_1, \cdots, Y_n) = a, \quad s_0(n; Y_1, \cdots, Y_n) = b\}, \tag{40}$$

using notation indicated at the beginning of Sec. 1. On the other hand, from the hypotheses of the lemma, (39) also equals

$$\Pr\left\{u(n; Y_1, \cdots, Y_n) \in S \mid t_0(n; Y_1, \cdots, Y_n) = \frac{a - \theta}{\sigma}, \quad s_0(n; Y_1, \cdots, Y_n) = \frac{b}{\sigma}\right\}. \tag{41}$$

From the equality of (40) and (41), it follows that the conditional distribution of u is independent of the conditioning values of t_0 and s_0, so that u is stochastically independent of (t_0, s_0). But t_0 and s_0 are mutually independent, since Corollary 2.2 applies. Hence the result.

This lemma can be used to compare the relative merits of s_0 and any other s asymptotically as $\sigma \rightarrow \infty$. Let us assume that the hypotheses of Lemma 4 are satisfied, that F is continuous and that t_0 also satisfies Assumptions III and V. Then we know that t_0 is the best statistic to use as "t", and both s_0 and s are eligible as the "s" statistic. Let

$$J(u) = \Pr\{s(m; X_1, \cdots, X_m) \leqq us_0(m; X_1, \cdots, X_m)\}, \tag{42}$$

and $H(u)$, $H_0(u)$ denote the c.d.f.'s of s and s_0 respectively. It will be understood that we have the same m throughout the discussion. We already know

$$\Pr\{t_0(n; X_1, \cdots, X_n) \leqq \theta + \sigma x\} = F\{xk_0(n)\}. \tag{43}$$

[2] My attention has been drawn to the fact that a general result of the type of those given in Lemmas 2 and 4 has been previously given, for boundedly complete sufficient statistics, by D. Basu [*Sankhya*, vol. 15 (1955), pp. 377–380.]

Let

$$M(y) = \int_0^\infty F(yu)\, dH_0(u). \tag{44}$$

Then

$$H(v) = \int_0^\infty H_0(v/u)\, dJ(u) \tag{45}$$

by Lemma 4, and we get

$$\int_0^\infty F(yu)\, dH(u) = \int_0^\infty M(yu)\, dJ(u). \tag{46}$$

From (5), (8), and (9) on account of continuity, we have

$$\rho = (\chi - \chi')/\delta, \qquad \rho_0 = (\chi_0 - \chi_0')/\delta, \tag{47}$$

and

$$\begin{cases} M(\chi_0) = 1 - \alpha = \int_0^\infty M(\chi u)\, dJ(u), \\ M(\chi_0') = 1 - \beta = \int_0^\infty M(\chi' u)\, dJ(u). \end{cases} \tag{48}$$

Hence,

$$(\chi_0 - \chi_0') = M^{-1}\left\{\int_0^\infty M(\chi u)\, dJ(u)\right\} - M^{-1}\left\{\int_0^\infty M(\chi' u)\, dJ(u)\right\}. \tag{49}$$

Now, from (14) and (15) we know that $E(N)$, $E(N_0) \to \infty$ as $\sigma \to \infty$, and

$$\frac{E(N_0)}{E(N)} \cong \frac{\int_0^\infty k_0^{-1}(\sigma\rho_0 u)\, dH_0(u)}{\int_0^\infty k^{-1}(\sigma\rho u)\, dH(u)}.$$

Suppose

$$k(u) = u^{1/c}, \tag{50}$$

where c is a constant $\geqq 1$. (This is the case in the normal and exponential populations.) Then

$$\frac{E(N_0)}{E(N)} \cong \frac{\rho_0^c \int_0^\infty u^c\, dH_0(u)}{\rho^c \int_0^\infty u^c\, dH(u)} = \frac{\rho_0^c}{\rho^c \int_0^\infty u^c\, dJ(u)}, \tag{51}$$

by (45).

Therefore

$$\rho\left\{\int_0^\infty u^c\, dJ(u)\right\}^{1/c} > \rho_0$$

implies that, asymptotically as $\sigma \to \infty$, $E(N_0) < E(N)$. However, since

$$\left\{\int_0^\infty u^c\, dJ(u)\right\}^{1/c} > \int_0^\infty u\, dJ(u),$$

if we can show that

$$(52) \qquad (\chi - \chi') \int_0^\infty u\, dJ(u) > \chi_0 - \chi_0',$$

this implies that asymptotically (t_0, s_0) is the best, or minimum-expected-sample-size, pair among those satisfying the initial assumptions.

We shall now prove (52) to hold in the two cases that matter. As previously noted, s_0 and as_0, where a is a constant >0, are equivalent statistics for our purpose, and hence in what follows we shall only consider as alternative candidates, statistics s which are not constant multiples of s_0; in other words, we assume that $J(u)$ has at least two points of increase.

Example 1. Let $F(x) = \int_{-\infty}^{x} e^{-(u^2/2)}\, du/\sqrt{2\pi}$, and assume $\alpha < 0.5 < \beta$. $M(y)$ is Student's distribution, so that $M(0) = 0.5$. Hence

$$(53) \qquad \chi_0' < 0 < \chi_0 \quad \text{and} \quad \chi' < 0 < \chi.$$

Further, $M(y)$ is concave or convex according as $y > 0$ or <0, and therefore $M^{-1} \int_0^\infty M(yu)\, dJ(u) \lessgtr y \int_0^\infty u\, dJ(u)$ according as $y \gtrless 0$. From (53) and (49), (52) follows.

Example 2. Let $F(x)$ be given by (17). Then

$$M(y) = \begin{cases} 0 & \text{if } y \leqq 0 \\ 1 - (1+y)^{-\mu} & \text{if } y > 0, \end{cases} \qquad \text{where } \mu = m - 1 \geqq 1.$$

Consequently, all χ's are positive. Now let

$$(54) \qquad \begin{aligned} f(y) &= y \int_0^\infty u\, dJ(u) - M^{-1} \int_0^\infty M(yu)\, dJ(u) \\ &= y \int_0^\infty u\, dJ(u) - \left\{\int_0^\infty (1 + yu)^{-\mu}\, dJ(u)\right\}^{-1/\mu} - 1. \end{aligned}$$

Then

$$f'(y) = \int_0^\infty u\, dJ(u) - \int_0^\infty u(1 + yu)^{-\mu-1}\, dJ(u) \cdot \left\{\int_0^\infty (1 + yu)^{-\mu}\, dJ(u)\right\}^{-(\mu+1)/\mu}$$

It is easily seen that $f'(y) > 0$ for $y > 0$; because the sign of $f'(y)$ is the same as that of

$$\left\{\int_0^\infty (1 + yu)^{-\mu} \, dJ(u)\right\}^{(\mu+1)/\mu} \int_0^\infty u \, dJ(u) - \int_0^\infty u(1 + yu)^{-\mu-1} \, dJ(u)$$

which is

$$> \left\{\int (1 + yu)^{-\mu} \, dJ(u)\right\}\left\{\int (1 + yu)^{-1} \, dJ(u)\right\} \int u dJ(u)$$

$$- \int u(1 + yu)^{-\mu-1} \, dJ(u)$$

$$> \int (1 + yu)^{-\mu} \, dJ(u) \int u(1 + yu)^{-1} \, dJ(u) - \int u(1 + yu)^{-\mu-1} \, dJ(u)$$

$$> 0,$$

since u and $(1 + yu)^{-1}$ are monotone in opposite directions for $y > 0$, and the same is true of $u(1 + yu)^{-1}$ and $(1 + yu)^{-\mu}$. Consequently, $f(y)$ is an increasing function of $y > 0$; (52) follows from (49), and asymptotically as $\sigma \to \infty$, (t_0, s_0) is the best pair of statistics to use.

REFERENCES

[1] C. Stein, "A two-sample test of a linear hypothesis whose power is independent of the variance," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 243–258.

[2] S. G. Ghurye and Herbert Robbins, "Two-stage procedures for estimating the difference between means," *Biometrika*, Vol. 41 (1954), pp. 146–152.

[3] E. J. G. Pitman, "The estimation of the location and scale parameters of a continuous population of any given form," *Biometrika*, Vol. 30 (1939), pp. 391–421.