We have thus shown that any test is dominated by a test in $\mathfrak{A}$, i.e. that $\mathfrak{A}$ is essentially complete. It remains to prove admissibility. Suppose ϕ and ϕ^* are given by g and g^*. Without changing the characteristics of the tests, we may redefine g and g^* so that they are left-continuous and so that $g(u) = -1$ where $g(u) \leqq u$, and $g^*(u) = -1$ where $g^*(u) \leqq u$. Suppose there is a u' such that $g(u') > g^*(u')$. Choose u'' such that $g(u') > u'' > g^*(u')$. (See the diagram.) Let "area" be measured with respect to the density $2^{-n}n(n-1)(v-u)^{n-2}du\,dv$. By left-continuity, $g^*(u) < u$ for all u in an interval whose right endpoint is u'. Therefore either the "area" below g in $T(u'+1)$ is less than that below g^*, or the "area" below g in $T(u''-1)$ is greater than that below g^*. But the "area" below g in $T(\theta)$ is just $E_\theta(\phi)$. Thus either $E_{u'+1}(\phi) < E_{u'+1}(\phi^*)$ or $E_{u''-1}(\phi) > E_{u''-1}(\phi^*)$. But $u' + 1 > 0$ and $u'' - 1 < 0$, so this shows ϕ doesn't dominate ϕ^*. Hence if ϕ dominates ϕ^*, $g(u') \leqq g^*(u')$ for all u'. But in this case either ϕ and ϕ^* are essentially the same or $E_\theta(\phi) < E_\theta(\phi^*)$ for sufficiently small positive θ. Therefore ϕ cannot dominate ϕ^*. Since ϕ and ϕ^* were arbitrary tests of the essentially complete class $\mathfrak{A}$, it follows that all tests in A are admissible.

This proof of admissibility is spelled out analytically in [2]. The proof of essential completeness given there uses a general property possessed by the rectangular distribution.

REFERENCES

[1] A. Birnbaum, "Admissible tests for the mean of a rectangular distribution," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 157–161.

[2] J. W. Pratt, "Some results in the decision theory of one-parameter multivariate Polya type distributions," Stanford Technical Report No. 37 (Oct. 21, 1955), pp. 45–50.

A METHOD FOR SELECTING THE SIZE OF THE INITIAL SAMPLE IN STEIN'S TWO SAMPLE PROCEDURE

By Jack Moshman

Corporation for Economic and Industrial Research, Arlington 2, Virginia

1. Summary and Introduction. The use of an upper percentage point of the distribution of total sample size in conjunction with the expectation of the latter is proposed as a guide to the selection of the size of the initial sample when using some version of Stein's [5] two-sample procedure. It is a rapidly calculable function of the underlying population variance based on existing tables of the χ^2 distribution. A rule-of-thumb is proposed to be used in making the actual selection of initial sample size. It is a simple matter to investigate the nature of the percentage point for different values of the variance over a limited range;

Received July 22, 1957; revised December 2, 1957.

a recommended conservative choice when the variance is not known is the selection of a large initial sample.

Dantzig [2] proved the nonexistence of nontrivial tests of Student's hypothesis whose power was independent of the variance, a result extended by Stein to the general linear hypothesis. In the same paper Stein proposed a two-sample procedure whose power was independent of variance. The same two-sample method could be used to obtain a confidence interval for the mean of a normal distribution with predetermined length and confidence coefficient.

Stein gave no specifications for the choice of the initial sample size, but Seelbinder [4] suggested that it be selected to minimize the expectation of the total sample. In a recent paper, Bechhofer, Dunnett and Sobel [1] used Stein's procedure for another application, noting that the variance of the total sample size increased as the size of the first sample decreased.

An efficient choice of the size of the initial sample will hold the expectation of the sample small, and will further reduce the probability of an extremely large total sample. This note will explore the matter in further detail and show that an upper percentage point of the distribution of total sample size, when used in conjunction with the expectation, is a rapidly calculable guide to an efficient choice of the size of the first sample.

2. Basic theory. As developed by Stein, the two-sample procedure involves a preliminary, arbitrary choice of a positive integer N_0 and a number $z > 0$. The value of z will depend, when constructing a confidence interval of length $2L$ for the mean, on the precision of the estimate, i.e., the length of the interval, and its reliability, the confidence coefficient. Specifically, if $t_{n,\gamma}$ is the upper 100 γ percentage point of Student's distribution with n degrees of freedom, one would take $z = L^2/t^2_{N_0-1,1-(\alpha/2)}$ to obtain a confidence coefficient $\geqq 1 - \alpha$.

A sample of N_0 observations is taken and $s^2 = \sum(x_i - \bar{x})^2/(N_0 - 1)$ is computed as an estimate of the unknown variance σ^2 with $n = N_0 - 1$ degrees of freedom. The total sample size, N, is then

$$N = \max\left(\left[\frac{s^2}{z}\right] + 1, N_0\right), \tag{1}$$

where $[t]$ is the *largest integer less than t*.

Hence it follows that

$$\text{Prob}\,(N = N_0) = \text{Prob}\left(\frac{s^2}{z} \leqq N_0\right) = \text{Prob}\left(\frac{ns^2}{\sigma^2} = \chi^2(n) \leqq \frac{nN_0\,z}{\sigma^2}\right), \tag{2}$$

where $\chi^2(n)$ is distributed as χ^2 with n degrees of freedom. Furthermore, for integral $m > N_0$,

$$\begin{aligned} \text{Prob}\,(N = m) &= \text{Prob}\left(m - 1 < \frac{s^2}{z} \leqq m\right) \\ &= \text{Prob}\left(\frac{n(m-1)z}{\sigma^2} < \chi^2(n) \leqq \frac{nmz}{\sigma^2}\right). \end{aligned} \tag{3}$$

Therefore, letting $\lambda = z/\sigma^2$, one may easily show

$$E(N) = N_0 \operatorname{Prob}(\chi^2(n) < n\lambda N_0) + \frac{1}{\lambda} \operatorname{Prob}(\chi^2(n+2) > n\lambda N_0) + \theta_1 \operatorname{Prob}(\chi^2(n) > n\lambda N_0) \tag{4}$$

and

$$\begin{aligned}\operatorname{Var}(N) = {} & N_0^2 \operatorname{Prob}(\chi^2(n) < n\lambda N_0) + \frac{(n+2)}{n\lambda^2} \operatorname{Prob}(\chi^2(n+4) > n\lambda N_0) \\ & + \frac{2\theta_2}{\lambda} \operatorname{Prob}(\chi^2(n+2) > n\lambda N_0) + \theta_3 \operatorname{Prob}(\chi^2(n) > n\lambda N_0) - (E(N)^2,\end{aligned} \tag{5}$$

where $0 \leqq \theta_i \leqq 1$, $i = 1, 2, 3$.

Whereas (4) defines $E(N)$ within a maximum error of unity, (5) is not as useful inasmuch as the factor $1/\lambda$ may be, and frequently is, large.

Furthermore, it is somewhat difficult to translate Var (N) into working percentage points of the distribution of N. A more useful procedure is to calculate a given percentage point N_p of the distribution. This may be accomplished directly from (2) and (3). Define N_p as the smallest integer $\geqq N_0$ such that

$$\operatorname{Prob}(N \leqq N_p) = \sum_{m=N_0}^{N_p} \operatorname{Prob}(N = m) \geqq p. \tag{6}$$

But this is equivalent, if one writes $p_n(\chi^2)$ as the probability density function of $\chi^2(n)$, to setting

$$\int_0^{nN_p\lambda} p_n(\chi^2)\, d\chi^2 \geqq p \tag{7}$$

and letting N_p be chosen to satisfy (7), but not less than N_0. Thus

$$N_p = \max\left\{N_0, \left[\frac{1}{\lambda}\left(100p\text{th } \textit{percentage point of } \frac{\chi^2(n)}{n}\right)\right] + 1\right\}, \tag{8}$$

which is tabulated in Hald [3] for example. Note that the upper percentage points of $\chi^2(n)/n$ decreases monotonically as n increases. Conceivably, if N_0 is chosen very large, one can be reasonably confident that no further sampling will be necessary, but this is not an efficient procedure.

A rough, but objective, rule-of-thumb may be derived by the following consideration: Let $E(N \mid N_0^*)$ be the expectation of N if $N_0 = N_0^*$ and $N_p(N_0^*)$ the 100pth percentile of N if $N_0 = N_0^*$. Define

$$P(N_0^*) = \int_0^{n\lambda E(N \mid N_0^*)} p_n(\chi^2)\, d\chi^2, \tag{9}$$

as the proportion of time N will not exceed $E(N \mid N_0^*)$. Let $\mathbf{N}_0$ be the value of N_0 which minimizes $E(N)$, i.e.,

$$E(N \mid \mathbf{N}_0) \leqq E(N \mid N_0) \tag{10}$$

for all N_0. Now one might investigate alternative values of N_0 by considering

$$\Psi(N_0) = (1 - p)(N_p(\mathbf{N}_0) - N_p(N_0)) - (1 - P(\mathbf{N}_0))(E(N \mid N_0) - E(N \mid \mathbf{N}_0)) \tag{11}$$

and selecting N_0 as the integer for which $\Psi(N_0)$ is a maximum. In effect, (11) weights the expected changes in $E(N)$ and N_p by the probability of exceeding those values. It would be expected that p would be chosen independently from nonstatistical considerations.

3. Example. If one takes $\lambda = .1$, where in the ordinary application considered by Stein $n = N_0 - 1$, then $E(N)$ is a minimum for $N_0 = \mathbf{N}_0 = 3$. Values of $E(N \mid N_0)$ are tabulated in Table 1. It may be seen that $E(N \mid N_0)$ is fairly constant over a considerable range. The same table also contains $N_{.95}(N_0)$ which decreases sharply where $E(N \mid N_0)$ is relatively constant.

It may readily be verified from (9) that $P(N_0) = P(3) \approx .64$. Rapidly one may evaluate $\Psi(N_0)$ from (11), taking $p = .95$, and find that $\Psi(6) = .2686$ is the maximum. Hence the rule suggested specifies $N_0 = 6$ as the proper choice.

4. Discussion. When the variance is unknown, two alternatives exist. It may be feasible to express the length of the confidence interval desired as a proportion of σ; no difficulty then ensues since λ is specified. If L is specified absolutely, in most practical cases a range for σ is known. One can then investigate the distribution of N for various values of σ in this range and make a subsequent choice of N_0.

The procedures suggested in this note are particularly applicable to those situations where repeated sampling is not contemplated and/or there exists a physical reason for wanting to avoid excessively large samples. The latter situation may obtain where larger individual samples may entail the purchase of additional test equipment or require the supplementing of a regular interviewing staff by extra employees.

TABLE 1

Dependence of $E(N)$ and $N_{.95}$ on N_0

$\lambda = .1$

N_0	$E(N \mid N_0)$	$N_{.95}(N_0)$	N_0	$E(N \mid N_0)$	$N_{.95}(N_0)$
2	10.45	38.41	10	11.84	18.80
3	10.29	29.96	12	12.92	17.89
4	10.45	26.05	14	14.35	17.20
5	10.51	23.72	16	16.15	16.66
6	10.63	22.14	18	18.02	18.00
7	10.80	20.99	20	20.02	20.00
8	11.18	20.10	22	22.01	22.00
9	11.43	19.38	24	24.00	24.00

5. Acknowledgment. The author wishes to express his indebtedness to Professor G. E. Albert for many helpful suggestions made in the pursuance of this research.

REFERENCES

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

[2] G. B. Dantzig, "On the non-existence of tests of 'Student's' hypothesis having power functions independent of σ," *Ann. Math. Stat.*, Vol 11 (1940), p. 186.

[3] Anders Hald, *Statistical Theory with Engineering Applications*, John Wiley & Sons, 1952.

[4] B. M. Seelbinder, "On Stein's two-stage sampling scheme," *Ann. Math. Stat.*, Vol. 24 (1953), p. 640.

[5] Charles Stein, "A two-sample test for a linear hypothesis whose power is independent of the variance," *Ann. Math. Stat.*, Vol. 16 (1945), p. 243.

ON A PROBLEM IN MEASURE-SPACES

By V. S. Varadarajan

Indian Statistical Institute, Calcutta

Summary. Let $\mathfrak{F}$ be the family of all random variables on a probability space Ω taking values from a separable and complete metric space X. In this paper we prove that $\mathfrak{F}$ is in a certain sense a closed family. More precisely, if $\{\xi_n\}$ is a sequence of X-valued random variables such that their probability distributions converge weakly to a probability distribution P on X, then there exists an X-valued random variable on Ω with distribution P. An example is also given which shows that the assumption of completeness of X cannot in general be dropped.

1. Preliminary remarks. In what follows $(\Omega, \mathcal{S}, \mu)$ is a probability space and X a separable metric space. We denote by $\mathcal{B}$ the class of Borel subsets of X defined as the minimal σ-field containing all open subsets of X.

A map φ of Ω into X is called a random variable if it is measurable i.e., $\varphi^{-1}(A) \varepsilon \mathcal{S}$ for each $A \varepsilon \mathcal{B}$. If φ is a random variable we define as its distribution the measure μ_φ on $\mathcal{B}$ given by

$$\mu_\varphi(A) = \mu\{\varphi^{-1}(A)\}$$

for all $A \varepsilon \mathcal{B}$. A given probability measure P on $\mathcal{B}$ is said to be induced from Ω if there exists a random variable φ such that $P = \mu_\varphi$.

Suppose we are given a sequence $\{P_n\}$ of probability measures on $\mathcal{B}$. We say that $\{P_n\}$ converges weakly to a probability measure P on $\mathcal{B}$ ($P_n \Rightarrow P$ in symbols) if

$$\lim_{n\to\infty} \int_X g\, dP_n = \int_X g\, dP$$

Received November 12, 1956; revised May 19, 1958.