NOTES

This section is devoted to brief research and expository articles on methodology and other short items.

ON THE EXPECTED VALUES OF TWO STATISTICS

By H. E. Robbins

Post Graduate School; Annapolis, Md.

In a previous paper[1], the following theorem was proved. Let X be a random, Lebesgue measurable subset of Euclidean m dimensional space E_m, and let $\mu(X)$ be the measure of X. For every point x of E_m let $p(x)$ be the probability that X contains x. Then

$$E(\mu(X)) = \int_{E_m} p(x)\, d\mu(x). \tag{1}$$

In the present note we shall show how this theorem may be used to find the expected values of two statistics which arise in sampling theory. Applications to similar problems may suggest themselves to the reader.

Let Y be a real random variable with c. d. f. (cumulative distribution function) $\sigma(y)$, so that for every y,

$$\Pr\ (Y < y) = \sigma(y). \tag{2}$$

Let $Y_1, \cdots, Y_n$ be n independent random variables, each with the distribution of Y. Finally, let

$$\begin{aligned} A &= \min\ (Y_1, \cdots, Y_n), \\ B &= \max\ (Y_1, \cdots, Y_n), \\ R &= B - A, \\ F &= \sigma(B) - \sigma(A). \end{aligned} \tag{3}$$

Although the values of $E(F)$ and $E(R)$ can be found from the sampling distributions of F and R, and, in fact, are well known, we shall show how to apply (1) to find $E(F)$ and $E(R)$ directly.

To find the first of these, let X denote the set of points in the interval $0 \leq x \leq 1$ such that

$$\sigma(A) < x < \sigma(B). \tag{4}$$

Then X is a random set with measure

$$\mu(X) = F. \tag{5}$$

Moreover, for any point x the probability that X shall contain x is clearly

$$p(x) = 1 - x^n - (1 - x)^n. \tag{6}$$

Hence by (1),

$$E(\mu(X)) = \int_0^1 [1 - x^n - (1 - x)^n]\, dx = \frac{n-1}{n+2}. \tag{7}$$

Thus by (5),

$$E(F) = \frac{n-1}{n+1}. \tag{8}$$

This result may also be derived by the usual method. In fact, it is not hard to show that the c. d. f. of F is

$$\tau(f) = \Pr (F < f) = (1 - n)f^n + nf^{n-1} \quad \text{for} \quad 0 \le f \le 1, \tag{9}$$

whence

$$\begin{aligned} E(F) &= \int_0^1 f\, d\tau(f) = (1 - n)n \int_0^1 f^n\, df + n(n-1) \int_0^1 f^{n-1}\, df \\ &= \frac{(1-n)n}{n+1} + \frac{n(n-1)}{n} = \frac{n-1}{n+1}. \end{aligned} \tag{10}$$

Here the advantage of using (1) is only that it makes unnecessary the calculation of the c. d. f. $\tau(f)$, provided that only $E(F)$ is desired.

The situation is quite otherwise with $E(R)$. Here the c. d. f. of R is

$$\theta(r) = \Pr (R < r) = n(n-1) \int_{-\infty}^{\infty} \varphi(a) \int_a^{a+r} \varphi(b) \left[\int_a^b \varphi(t)\, dt\right]^{n-2} db\, da, \tag{11}$$

where φ is the probability density function of Y. Unless φ is a very simple function, it would seem difficult to find a simple expression for the integral

$$E(R) = \int_0^{\infty} r\, d\theta(r). \tag{12}$$

However, if we let X now denote the linear set

$$A \le t \le B, \tag{13}$$

then

$$\mu(X) = B - A = R. \tag{14}$$

The probability that X shall contain the point t is now

$$p(t) = 1 - \sigma^n(t) - (1 - \sigma(t))^n, \tag{15}$$

so that, by (1) and (14),

$$E(R) = \int_{-\infty}^{\infty} \{1 - \sigma^n(t) - (1 - \sigma(t))^n\}\, dt. \tag{16}$$

This formula for the expected value of the range in a sample of n from a population Y with c. d. f. $\sigma(t)$ is believed to be new.

If $\sigma(t)$ is such that $dt/d\sigma$ can be found as an explicit function of σ, then (16) can be written with advantage as

$$E(R) = \int_0^1 \{1 - \sigma^n - (1 - \sigma)^n\} \frac{dt}{d\sigma} \, d\sigma. \tag{17}$$

For example, suppose the random variable Y has the probability density function

$$\varphi(y) = \frac{e^y}{(1 + e^y)^2}, \tag{18}$$

and hence the c. d. f.

$$\sigma(y) = \frac{e^y}{1 + e^y}. \tag{19}$$

Then

$$t = \log \frac{\sigma}{1 - \sigma}, \qquad \frac{dt}{d\sigma} = \frac{1}{\sigma(1 - \sigma)}. \tag{20}$$

Hence from (17), the expected value of the range in a sample of n is

$$E(R) = \int_0^1 \frac{1 - \sigma^n - (1 - \sigma)^n}{\sigma(1 - \sigma)} \, d\sigma. \tag{21}$$

The indicated division in the integrand may be carried out, and the result, a polynomial in σ of degree $\leq (n - 2)$, when integrated between 0 and 1, gives an explicit formula for $E(R)$. Thus for samples of $n = 2, 3, 4$ we find the values of $E(R)$ to be respectively 2, 3, 11/3. Incidentally, it is always true that the expected value of the range for $n = 3$ is three-halves that for $n = 2$. This follows from (16) and the algebraic identity

$$\{1 - \sigma^3 - (1 - \sigma)^3\} = \tfrac{3}{2}\{1 - \sigma^2 - (1 - \sigma)^2\}. \tag{22}$$

REFERENCES

[1] H. E. Robbins, "On the measure of a random set," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 70–74.

ON RELATIVE ERRORS IN SYSTEMS OF LINEAR EQUATIONS

By A. T. Lonseth

Northwestern University

Some time ago in these *Annals*[1], L. B. Tuckerman discussed the effect of relative coefficient errors on relative solution errors for a non-singular linear algebraic system; his discussion was confined to errors so small that their squares and higher powers can be neglected. Dr. Tuckerman's paper was principally concerned

[1] L. B. Tuckerman, "On the mathematically significant figures in the solution of simultaneous linear equations," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 307–316.