JOINT SAMPLING DISTRIBUTION OF THE MEAN AND STANDARD DEVIATION FOR PROBABILITY DENSITY FUNCTIONS OF DOUBLY INFINITE RANGE[1]

By Melvin D. Springer

U. S. Naval Ordnance, Indianapolis

1. Summary. The joint sampling distribution of $\bar{x}$ and S is derived in integral form for probability density functions of doubly infinite range. This derivation is effected through the use of a transformation which transforms the sample probability element $f(x_1)f(x_2) \cdots f(x_n)\, dx_1\, dx_2 \cdots dx_n$ into the element

$$f(x_1)f(x_2) \cdots f(x_{n-2})f((n\bar{x} - \textstyle\sum_1^{n-2} x_i \pm \Omega_1)/2)f((n\bar{x} - \sum_1^{n-2} x_i \mp \Omega_1)/2) \mid J \mid \cdot dx_1\, dx_2 \cdots dx_{n-2}\, d\bar{x}\, dS,$$

where $\bar{x} = (1/n) \sum_1^n x_i$, $S^2 = (1/n) \sum_1^n (x_i - \bar{x})^2$, and J is the Jacobian of the transformation. Bounds on x_{n-r}, $r = 2, 3, \cdots, n - 1$, are established in terms of $\bar{x}$, S, and x_{n-r-j}, $j = 1, 2, \cdots, n - r - 1$. The probability element

$$f(x_1)f(x_2) \cdots f(x_{n-2})f((n\bar{x} - \textstyle\sum_1^{n-2} x_i \pm \Omega_1)/2)f((n\bar{x} - \sum_1^{n-2} x_i \mp \Omega_1)/2) \mid J \mid dx_1\, dx_2 \cdots dx_{n-2}\, d\bar{x}\, dS$$

must then be integrated with respect to x_{n-r}, $r = 2, 3, \cdots, n - 1$, between these limits to obtain $F(\bar{x}, S)\, d\bar{x}\, dS$, the joint probability element of $\bar{x}$ and S. These limits of integration of x_{n-r}, $r = 2, 3, \cdots, n - 1$ enable one to express $F(\bar{x}, S)$ in terms of quadratures when $f(x)$ is any probability density function of doubly infinite range. To illustrate the method, $F(\bar{x}, S)$ is obtained when $f(x)$ is the normal probability density function.

2. Introduction. It is well known that if random samples of n items are drawn from a parent population, $\bar{x}$ and S will be independent in the probability sense if and only if x is normally distributed in this population [1]. Furthermore, if the parent population is normal with mean m and standard deviation σ, $\bar{x}$ and S are distributed jointly in accordance with

$$F(\bar{x}, S) = \frac{n^{n/2}S^{n-2} \exp\{-(n/2\sigma^2)[(\bar{x} - m)^2 + S^2]\}}{\sigma^n 2^{((n-2)/2)}\pi^{\frac{1}{2}}\Gamma\left(\frac{n-1}{2}\right)}. \tag{1}$$

Received 6/2/51, revised 8/28/52.

[1] This paper was presented at a joint meeting of the Institute of Mathematical Statistics and the Biometric Society at Oak Ridge, Tennessee on March 17, 1951. The opinions expressed herein are solely those of the author and are not necessarily those of the U. S. Navy Department.

This joint distribution for the normal function is often referred to as Helmert's distribution, since it was first established by Helmert [2]. Helmert arrived at (1) through the use of a pair of linear transformations which transformed the joint distribution of the individual errors of observation into a joint distribution of sample mean error and standard deviation, plus dummy variables which were integrated out over all possible values. Kruskal [3] has shown that Helmert's distribution may be obtained directly by mathematical induction. However, when sampling is extended to nonnormal universes, little seems to be known about $F(\bar{x}, S)$ except for very small samples. Truksa [4] has expressed $F(\bar{x}_3, S_3)$ in integral form and has applied "the concept of the probability of passage" to obtain $F(\bar{x}_{t+2}, S_{t+2})$ from $F(\bar{x}_t, S_t)$, where $\bar{x}_t$ and S_t represent, respectively, the mean and standard deviation of a sample of t items and where $F(\bar{x}_t, S_t)$ is assumed to be known. A. T. Craig [5] has derived $F(\bar{x}, S)$ in integral form *when* $n = 2, 3, 4$ for probability density functions of doubly infinite, singly infinite, and finite positive range. Yet, for no probability density function $f(x)$ has $F(\bar{x}, S)$ ever been expressed explicitly in terms of quadratures for the general case of samples of size n. It is the purpose of this paper to derive $F(\bar{x}, S)$ in terms of quadratures *for any sample size* when $f(x)$ is any probability density function of doubly infinite range. Whereas the procedure in [3] and [4] is to add one or two new observations and express the new $\bar{x}$ and S in terms of the old, *whose distribution is taken as known*, I shall employ a transformation *for fixed* n and derive an integration formula, particularly the limits of integration, inductively.

3. $F(\bar{x}, S)$ for probability density functions of doubly infinite range. Consider a universe characterized by the probability density function $f(x)$, $-\infty < x < \infty$. If n variates x_i, $i = 1, 2, \cdots, n$, are selected at random from this universe the probability that they will fall simultaneously within the intervals dx_i, $i = 1, 2, \cdots, n$, is given, to within infinitesimals of higher order, by

$$f(x_1)f(x_2) \cdots f(x_n)\, dx_1\, dx_2 \cdots dx_n .$$

Since $nS^2 = \sum_1^n x_i^2 - n\bar{x}^2$ and $n\bar{x} = \sum_1^n x_i$, we may eliminate x_n in the first equation, obtaining

$$\sum_1^{n-1} x_i^2 + u_{n-1}^2 - 2n\bar{x}u_{n-1} + n(n-1)\bar{x}^2 - nS^2 = 0, \tag{2}$$

where $u_k = \sum_1^k x_i$. Solving this (symmetric) equation for x_{n-1} we have

$$x_{n-1} = \tfrac{1}{2}(n\bar{x} - u_{n-2} \pm \Omega_1), \tag{3}$$

where

$$\begin{aligned}\Omega_r^2 = -r(r+1) \sum_1^{n-r-1} x_i^2 - ru_{n-r-1}^2 + 2rn\bar{x}u_{n-r-1} \\ - rn(n-r-1)\bar{x}^2 + (r+1)rnS^2.\end{aligned} \tag{4}$$

Thus we may employ the transformation T:

$$\begin{aligned} x_i &= x_i, & i = 1, 2, \cdots, n-2, \\ x_{n-1} &= \tfrac{1}{2}(n\bar{x} - u_{n-2} \pm \Omega_1), \\ x_n &= \tfrac{1}{2}(n\bar{x} - u_{n-2} \mp \Omega_1). \end{aligned}$$

Application of this transformation to $f(x_1)f(x_2) \cdots f(x_n)\, dx_1\, dx_2 \cdots dx_n$ gives

$$\begin{aligned} &f(x_1)f(x_2) \cdots f(x_n)\, dx_1\, dx_2 \cdots dx_n \\ &= f(x_1)f(x_2) \cdots f(x_{n-2})f(\tfrac{1}{2}(n\bar{x} - u_{n-2} \pm \Omega_1))f(\tfrac{1}{2}(n\bar{x} - u_{n-2} \mp \Omega_1)) \\ &\qquad \cdot \mid J \mid dx_1\, dx_2 \cdots dx_{n-2}\, d\bar{x}\, dS, \end{aligned} \tag{5}$$

where $\mid J \mid = \mid$ Jacobian of $T \mid = n^2S/\Omega_1$. Evaluation of the multiple integral

$$\begin{aligned} &\int \cdots \iint f(x_1) \cdots f(x_{n-3})f(x_{n-2})f(\tfrac{1}{2}(n\bar{x} - u_{n-2} - \Omega_1)) \\ &\qquad \times f(\tfrac{1}{2}(n\bar{x} - u_{n-2} + \Omega_1))2n^2S/\Omega_1\, dx_{n-2}\, dx_{n-3} \cdots dx_1 \end{aligned} \tag{6}$$

over the range of the variables x_i, $i = 1, 2, \cdots, n-2$, yields the joint distribution $F(\bar{x}, S)$. It will be shown presently that the limits of integration of x_{n-r} in (6) are $(n\bar{x} - u_{n-r-1} \pm \Omega_r)/(r+1)$, $r = 2, 3, \cdots, n-1$. Before establishing these limits, let us consider an example.

4. The normal distribution. To illustrate the method, we shall derive $F(\bar{x}, S)$ for the normal distribution with mean m and standard deviation σ. This entails evaluating (6) for $f(x) = (1/\sigma(2\pi)^{\frac{1}{2}}) \exp\{-\frac{1}{2}(x-m)^2/\sigma^2\}$, the limits of integration of x_{n-r}, $r = 2, 3, \cdots, n-1$, having been specified at the close of Section 3. Upon employing the relationship (which is easily verified)

$$\Omega_m^2 = \frac{m}{m+2}\,\Omega_{m+1}^2 - m(m+2)\left\{x_{n-m-1} + \frac{1}{m+2}(u_{n-m-2} - n\bar{x})\right\}^2$$

and evaluating a few of the integrals in (6), it becomes evident that after r integrations we have

$$\begin{aligned} F(\bar{x}, S) &= \frac{4n^2S\exp\{-\frac{1}{2}(n/\sigma^2)[(\bar{x}-m)^2 + S^2]\pi^{r/2}\Gamma(\frac{3}{2})}{\sigma^n(2\pi)^{n/2}(r+1)^{(r-1)/2}(r+2)^{r/2}\Gamma\left(\frac{r+1}{2}\right)} \\ &\qquad \times \iint \cdots \int \Omega_{r+1}^{r-1}\, dx_{n-r-2} \cdots dx_2\, dx_1, \end{aligned} \tag{7}$$

the limits of integration having already been stated. To establish (7) by mathematical induction, assume that (7) results after r integrations are performed in (6), where r is any integer from 1 through $n-3$. Carrying out the next integration, we obtain in a very straightforward manner (7) with r replaced by $r+1$.

Thus, if (7) holds for r integrations, it necessarily holds for $r + 1$ integrations. But it is easily verified that (7) holds for $r = 1$; therefore, it holds when $r = 2, 3, \cdots, n - 2$. Letting $r = n - 2$ in (7), we have the well known joint sampling distribution of $\bar{x}$ and S for the normal universe, namely, (1).

5. Limits of integration of the variables. It remains to prove that the variable x_{n-r} is restricted to the closed interval

$$\left(\frac{n\bar{x} - u_{n-r-1} - \Omega_r}{r + 1}, \quad \frac{n\bar{x} - u_{n-r-1} + \Omega_r}{r + 1}\right), \qquad r = 2, 3, \cdots, n - 1. \tag{9}$$

To accomplish this, we again resort to mathematical induction. To expedite matters further, let us agree that when Ω_m^2, $m = 1, 2, \cdots, n - 2$, is involved in this discussion, it shall be regarded as a quadratic function of x_{n-m-1}. Bearing this in mind, we note that the discriminant of Ω_m^2 is Ω_{m+1}^2. We note further that since x_{n-1} is necessarily real, the inequality

$$\Omega_1^2 \geqq 0 \tag{10}$$

must be satisfied. Clearly, a necessary and sufficient condition that (10) be satisfied is that the discriminant of Ω_1^2 be nonnegative. That is, for a given $\bar{x}$ and S, the x_j, $j = 1, 2, \cdots, n - 3$, must collectively satisfy the condition

$$\Omega_2^2 \geqq 0, \tag{11}$$

in which case condition (10) will be fulfilled if and only if

$$\frac{n\bar{x} - u_{n-3} - \Omega_2}{3} \leqq x_{n-2} \leqq \frac{n\bar{x} - u_{n-3} + \Omega_2}{3}.$$

Similarly, condition (11) is met if and only if $\Omega_3^2 \geqq 0$. This restricts x_j, $j = 1, 2, \cdots, n - 4$, to values which collectively satisfy $\Omega_3^2 \geqq 0$, in which case

$$\frac{n\bar{x} - u_{n-4} - \Omega_3}{4} \leqq x_{n-3} \leqq \frac{n\bar{x} - u_{n-4} + \Omega_3}{4}.$$

In general, since the discriminant of Ω_{r-1}^2 is Ω_r^2, if $x_1, x_2, \cdots, x_{n-r}$ satisfy the condition

$$\Omega_{r-1}^2 \geqq 0, \qquad r = 2, 3, \cdots, n - 1, \tag{12}$$

$x_1, x_2, \cdots, x_{n-r-1}$ must necessarily fulfill, collectively, the condition $\Omega_r^2 \geqq 0$, whence

$$\frac{n\bar{x} - u_{n-r-1} - \Omega_r}{r + 1} \leqq x_{n-r} \leqq \frac{n\bar{x} - u_{n-r-1} + \Omega_r}{r + 1}.$$

That is, if condition (12) obtains when $r = \rho$, it necessarily holds when $r = \rho + 1$. But condition (12) must hold when $r = 2$; therefore, it must hold when $r = 3, 4, \cdots, n - 1$. This confines x_{n-r}, $r = 2, 3, \cdots, n - 1$, to the closed interval (9). Finally, since $\Omega_{n-1}^2 = n^2(n - 1)S^2 \geqq 0$, all the intervals (9) exist in the real domain.

Although the joint sampling distribution of $\bar{x}$ and S is given by (6) for any probability density function of doubly infinite range, it may be necessary to resort to numerical integration or other approximate methods to evaluate the multiple integral (6) when $n > 3$.

The distributions of $\bar{x}$ and S taken singly are, of course,

$$g(\bar{x}) = \int_0^\infty F(\bar{x}, S)\, dS, \tag{13}$$

and

$$h(S) = \int_{-\infty}^\infty F(\bar{x}, S)\, d\bar{x}. \tag{14}$$

The question naturally arises as to whether a similar method may not be used to determine the joint sampling distribution of $\bar{x}$ and S for probability density functions of singly infinite range. Actually, the procedure here described may be modified, particularly with respect to the limits of integration of (6), to obtain $F(\bar{x}, S)$ for probability density functions of singly infinite range. This modification, necessitated by the restriction of x_i, $i = 1, 2, \cdots, n$, to non-negative values, considerably complicates the derivation of $F(\bar{x}, S)$. Since the results are quite lengthy they will not be presented here, but will be discussed in detail in a later paper.

REFERENCES

[1] R. C. Geary, "The distribution of 'Student's' ratio for nonnormal samples," *J. Roy. Stat. Soc. Suppl.*, Vol. 3 (1936), pp. 178–184.

[2] F. R. Helmert, "Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungsfehlers direkter Beobachtungen gleicher Genauigkeit," *Astronomische Nachrichten*, Vol. 88 (1876), No. 2096, columns 113–132.

[3] William Kruskal, "Helmert's distribution," *Amer. Math. Monthly*, Vol. 53 (1946), pp. 435–438.

[4] L. Truksa, "The simultaneous distribution in samples of mean and standard deviation, and of mean and variance," *Biometrika*, Vol. 31 (1940), pp. 256–271.

[5] A. T. Craig, "The simultaneous distribution of mean and standard deviation in small samples," *Ann. Math. Stat.*, Vol. 3 (1932), pp. 126–140.