ON A CERTAIN DISTRIBUTION IN THE THEORY OF SAMPLING.

By K. V. KRISHNA SASTRY, *M.A., Mysore University, Mysore (S. India).*

(Communicated by Prof. K. B. Madhava, M.A., F.N.I.)

(*Received September 9; read November 22, 1946.*)

1. From a very large normal population with mean zero and standard deviation σ, i samples containing $n_1, n_2, \ldots . n_i$ individuals with means $m_1, m_2, \ldots . m_i$ and variances $s_1^2, s_2^2, \ldots . s_i^2$ are drawn. Under certain conditions Harold Simpson has shown that the number of times in which the numerical value of

$$\xi = \frac{a_1m_1+a_2m_2+\ \ldots\ +a_im_i}{\sqrt{(n_1s_1^2+n_2s_2^2+\ \ldots\ +n_is_i^2)}} \qquad \ldots \qquad \ldots \qquad \ldots \qquad (1)$$

exceeds $\tan\alpha \left(0 \leqslant \alpha \leqslant \frac{\pi}{2}\right)$ divided by the whole number of drawings has the value

$$\int_\alpha^{\frac{\pi}{2}} \cos^{t-1}\theta\, d\theta \Bigg/ \int_0^{\frac{\pi}{2}} \cos^{t-1}\theta\, d\theta$$

$t = (n_1+n_2+\ \ldots\ +n_i-i)$ and $a_1, a_2 \ldots a_i$ are constants.

Evidently $n_rs_r^2 = \Sigma(x-m_r)^2$ for all $r = 1, 2, \ldots i$ the summation sign extending over the n_r values of the sample.

In the present article the exact distribution of ξ has been derived without any restriction on the constants $a_1, a_2, \ldots a_i$ and H. Simpson's theorem is deduced as a special case.

2. Let $Z = a_1m_1+a_2m_2+\ \ldots .\ +a_im_i$

Then Z is normally distributed with mean zero and standard deviation S, where

$$S^2 = a_1^2\sigma_{m_1}^2+a_2^2\sigma_{m_2}^2+\ \ldots\ +a_i^2\sigma_{m_i}^2$$

$$= \sigma^2\left(\frac{a_1^2}{n_1}+\frac{a_2^2}{n_2}+\ \ldots\ +\frac{a_i^2}{n_i}\right) \qquad \ldots \qquad \ldots \qquad \ldots \qquad (2)$$

Hence

$$dF(z) = \frac{1}{S\sqrt{2\pi}}\, e^{-\frac{z^2}{2S^2}}\, dz \qquad \ldots \qquad \ldots \qquad \ldots \qquad (3)$$

Let $$u_r = n_rs_r^2 = \Sigma(x-m_r)^2.$$

Then the distribution of u_r is a type III distribution given by

$$dF(u_r) = \frac{1}{\left(\sigma\sqrt{2}\right)^{n_r-1}\Gamma\left(\frac{n_r-1}{2}\right)}\, e^{-\frac{u_r}{2\sigma^2}}\, u_r^{\frac{n_r-3}{2}}\, du_r \qquad \ldots \qquad \ldots \qquad (4)$$

The characteristic Function (C.F.) of this distribution is

$$f_r(t) = \frac{1}{(1-2i\sigma^2t)^{\frac{n_r-1}{2}}} \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots \qquad (5)$$

Let
$$u = n_1 s_1^2 + n_2 s_2^2 + \ldots + n_i s_i^2$$
$$= u_1 + u_2 + \ldots + u_i.$$

Since the samples are independent, the C.F. of u is given by
$$\phi(t) = f_1(t) \times f_2(t) \times \ldots \times f_i(t)$$
$$= \frac{1}{(1-2i\sigma^2 t)^k} \qquad \ldots \qquad (6)$$

Where $K = \frac{1}{2}(n_1 + n_2 + \ldots + n_i - i)$.
Hence the distribution function of u is also type III and is given by

$$dF(u) = \frac{1}{(2\sigma^2)^k \Gamma(k)} e^{-\frac{u}{2\sigma^2}} u^{k-1}\, du \qquad \ldots \qquad (7)$$

Now the problem reduces to finding the distribution of the ratio

$$\xi = \frac{z}{\sqrt{u}}$$

z is normally distributed and u is distributed as a type III.
It is shown by H. Cramer that this distribution is given by

$$dF(\xi) = \frac{\sigma}{S\sqrt{\pi}} \frac{\Gamma(K+\frac{1}{2})}{\Gamma(K)} \left(1 + \frac{\sigma^2 \xi^2}{S^2}\right)^{-K-\frac{1}{2}} d\xi \qquad \ldots \qquad (8)$$

which is the required distribution.
This is Pearsonian type VII distribution.
(3) If $a_1, a_2, \ldots a_i$ are chosen such that

$$\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2} + \ldots + \frac{a_i^2}{n_i} = 1$$

we have from (2), $S = \sigma$.
Then the distribution (8) becomes

$$dF = \frac{1}{\sqrt{\pi}} \frac{\Gamma(K+\frac{1}{2})}{\Gamma(K)} (1+\xi^2)^{-K-\frac{1}{2}} d\xi \qquad \ldots \qquad (9)$$

Now the probability that

$$|\xi| > \tan\alpha, \left(0 \leqslant \alpha \leqslant \frac{\pi}{2}\right)$$

is
$$P(|\xi| > \tan\alpha) = \int_{\tan\alpha}^{\infty} (1+\xi^2)^{-K-\frac{1}{2}} d\xi \Bigg/ \int_0^{\infty} (1+\xi^2)^{-K-\frac{1}{2}} d\xi.$$

Putting $\xi = \tan\theta$ and $2k = t$, we get

$$P(|\xi| > \tan\alpha) = \int_{\alpha}^{\frac{\pi}{2}} \cos^{t-1}\theta d\theta \Bigg/ \int_0^{\frac{\pi}{2}} \cos^{t-1}\theta\, d\theta \qquad \ldots \qquad (10)$$

which is Simpson's result.

References.

Cramer, Harold. Random Variables and Probability Distributions. *Camb. Tract.*, **36**.
Simpson, Harold (1943). On a Theorem Concerning Sampling. *J.R.S.S.*, pp. 266-7.
Uspeusky, J. V. Introduction to Mathematical Probability.