NOTES

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

THE DISTRIBUTION OF "STUDENT'S" RATIO FOR SAMPLES OF TWO ITEMS DRAWN FROM NON-NORMAL UNIVERSES

By Jack Laderman

The fundamental assumption in the derivation of "Student's" distribution[1]

$$f(t)\,dt = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}\,\Gamma\left(\frac{n-1}{2}\right)\left(1+\frac{t^2}{n-1}\right)^{\frac{1}{2}n}}\,dt$$

is that the universe sampled is normal. When the universe sampled is non-normal and n is small, the distribution of t differs considerably from "student's" distribution. In 1929, Rider[2] derived the distribution of t for samples of two drawn from the rectangular distribution

$$f(x)\,dx = \begin{cases} 1\,dx & \text{for } 0 \leq x \leq 1 \\ 0 & \text{elsewhere.} \end{cases}$$

In this paper, the formal expression for the distribution of t will be derived for samples of two drawn from any population having a continuous frequency function. A geometrical method similar to that employed by Rider will be used.

Let the universe sampled have the frequency function, $f(x)$, with zero mean, and let $f(x)$ be greater than zero from $x = a$ to $x = b$ and equal to zero elsewhere. Suppose the two observations are x_1 and x_2.

Then
$$\bar{x} = \frac{x_1 + x_2}{2}$$

and
$$t = \frac{\sqrt{n}\,\bar{x}}{\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}} = \frac{\sqrt{2}\,\bar{x}}{\sqrt{(x_1-\bar{x})^2 + (x_2-\bar{x})^2}}.$$

[1] "Student", "The Probable Error of a Mean" *Biometrika*, Vol. VI (1908), pp. 1–25.

[2] Paul R. Rider, "On the Distribution of the Ratio of the Mean to Standard Deviation in Small Samples from Non-Normal Universes", *Biometrika*, Vol. XXI (1929), pp. 124–143.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

The sample (x_1, x_2) can be represented by a point in a square of side $b-a$, as point P in Figure 1.

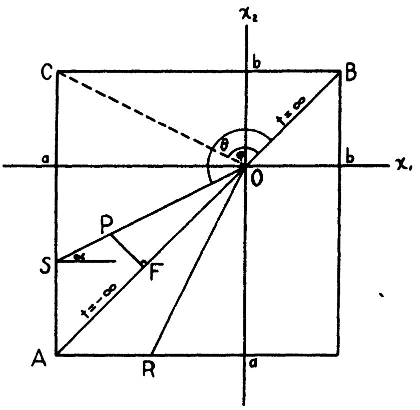

FIG. 1

The coordinates of F are $(\bar{x}, \bar{x})$

$$OF = -\sqrt{2}\,\bar{x}$$

$$FP = \sqrt{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2}$$

therefore
$$t = -\frac{OF}{FP} = \cot\theta.$$

Similarly for a point lying below AB, the value of t is $-\cot\theta$. Hence all points on OS and its image OR have the same value of t.

Let
$$\alpha = \theta - \frac{3\pi}{4}$$

Then
$$\tan\alpha = \frac{t+1}{t-1} \text{ and the equation of } OS \text{ is}$$

$$x_2 = \frac{t+1}{t-1}\,x_1.$$

The probability of getting a sample point in the element of area $dx_1\,dx_2$ is $f(x_1)f(x_2)\,dx_1\,dx_2$. Therefore the probability of getting a value of t less than the value represented by a point on OS is given by

$$2\int_a^0 \int_{x_1}^{\frac{t+1}{t-1}x_1} f(x_1)f(x_2)\,dx_2\,dx_1. \tag{1}$$

By differentiating (1), we get the frequency function

$$g(t) = -\frac{4}{(t-1)^2}\int_a^0 x_1 f(x_1) f\left(\frac{t+1}{t-1}\,x_1\right)dx_1.$$

However, this expression is valid only when $t \leq \cot \varphi$ where φ is the angle between OB and OC in Figure 1. From Figure 1 we notice that

$$\varphi = \frac{\pi}{4} + \cot^{-1}\left(-\frac{b}{a}\right)$$

$$\text{therefore } \cot \varphi = \frac{b+a}{b-a}.$$

Hence the above expression, $g(t)$, is valid for $t \leq \dfrac{b+a}{b-a}$.

When $t \geq \dfrac{b+a}{b-a}$, the probability of obtaining a value of t greater than the

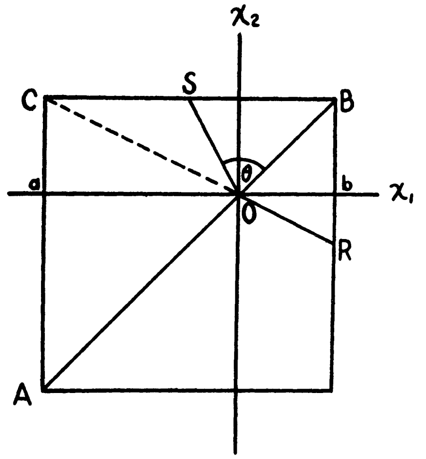

FIG. 2

value represented by a point on OS or its image OR, as in Figure 2, is given by

$$2 \int_0^b \int_{\frac{t-1}{t+1}x_2}^{x_2} f(x_1)f(x_2)\, dx_1\, dx_2$$

and the distribution function is

$$1 - 2 \int_0^b \int_{\frac{t-1}{t+1}x_2}^{x_2} f(x_1)f(x_2)\, dx_1\, dx_2. \tag{2}$$

After differentiating (2), we obtain the frequency function

$$g(t) = \frac{4}{(t+1)^2} \int_0^b x_2 f(x_2) f\left(\frac{t-1}{t+1}\, x_2\right) dx_2.$$

Thus, the frequency function of t for samples of two can be obtained from the expressions:

$$(3) \qquad g(t) = \begin{cases} \dfrac{4}{(t-1)^2} \displaystyle\int_0^a xf(x)f\left(\frac{t+1}{t-1}\,x\right)dx & \text{for } t \le \dfrac{b+a}{b-a} \\ \dfrac{4}{(t+1)^2} \displaystyle\int_0^b xf(x)f\left(\frac{t-1}{t+1}\,x\right)dx & \text{for } t \ge \dfrac{b+a}{b-a} \end{cases}$$

These expressions may also be used when a, or b, or both are infinite. However, the join point $\frac{b+a}{b-a}$ is then indeterminate, but by consideration of Figure 1, it can easily be seen that the join points are as follows:

$a = -\infty$ b finite	$t = -1$
a finite $b = +\infty$	$t = 1$
$a = -\infty$ $b = +\infty$	$t = 0$

The expressions given by (3) have been verified by obtaining the distribution of t for samples of two drawn from the normal distribution and also from the rectangular distribution. The explicit distributions were found very easily from (3) by performing the integrations.

For instance when $$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{x^2}{2\sigma^2}}$$

we get $$g(t) = \frac{1}{\pi(1+t^2)} \qquad -\infty < t < +\infty$$

which agrees with Student's distribution for $n = 2$.

And when $$f(x) = \begin{cases} 1 & -\frac{1}{2} \le x \le \frac{1}{2} \\ 0 & \text{elsewhere} \end{cases}$$

we get $$g(t) = \begin{cases} \dfrac{1}{2(t-1)^2} & \text{for } t \le 0 \\ \dfrac{1}{2(t+1)^2} & \text{for } t \ge 0 \end{cases}$$

which agrees with the distribution found by Rider as corrected by Perlo.[3]

NEW YORK

[3] Victor Perlo, "On the Distribution of Student's Ratio for Samples of Three Drawn from a Rectangular Distribution," *Biometrika*, Vol. XXV (1933) pp. 203–204.