On the Reduction Formulae for the Incomplete Probability Integral of the Studentised "D^2" Statistic

The distribution of the Studentised D^2[1,2] statistic in its earlier form was obtained by one of the authors, jointly with Mr. R. C. Bose as

$$\text{Const.}\ \frac{(D^2)^{\frac{p-2}{2}}}{(N+C^2D^2)^{\frac{N-1}{2}}}$$

$$ {}_1F_1\left(\frac{N-1}{2}; \frac{p}{2}; \frac{C^4\Delta^2}{2}\ \frac{D^2}{N+C^2D^2}\right) d(D^2) \qquad (1)$$

where ${}_1F_1$ is a certain type of hypergeometric function in Pochhammer's notation.

p = number of characters,

N = sum of sample sizes,

Δ^2 is the population parameter corresponding to D^2

and $C^2 = \frac{\bar{n}p}{2}$, $\bar{n}$ being the harmonic mean of the sample sizes.

LETTERS TO THE EDITOR

In its later generalised form where the mean differences are based on two samples and the pooled variances and co-variances are based on k samples, the distribution of Studentised D^2 was obtained by one of the authors jointly with Mr R. C. Bose[5] as

$$\text{Const.} \frac{(D^2)^{\frac{p-2}{2}}}{(N+C^2D^2)^{\frac{N-k+1}{2}}}$$
$$_1F_1\left(\frac{N-k+1}{2}, \frac{p}{2}, \frac{C^4\Delta^2}{2} \frac{D^2}{N+C^2D^2}\right) d(D^2) \quad (2)$$

where k = number of samples,

p and C^2 have the same meaning as before,

and N = sum of sizes of all the samples.

In both (1) and (2) the pooled variances and co-variances which enter into the construction of D^2 were obtained from the pooled (corrected) sum of squares by dividing by N, but if we adopt the modern convention of obtaining the variances from sum of squares by dividing by the appropriate number of degrees of freedom, which is N-2 for (1) and N-k for (2) then the distribution (2) which is really a generalisation of (1) takes on the form:

$$\text{Const.} \frac{(D^2)^{\frac{p-2}{2}}}{(N-k+C^2D^2)^{\frac{N-k+1}{2}}}$$
$$_1F_1\left(\frac{N-k+1}{2}, \frac{p}{2}, \frac{C^4\Delta^2}{2} \frac{D^2}{N-k+C^2D^2}\right) d(D^2) \quad (3)$$

To render this distribution useful for statistical purposes we have of course to find out the incomplete probability integral of D^2 which means the integral of (3) from 0 to D^2, since D^2 varies from 0 to ∞.

To facilitate this we put N-k = n. $\frac{C^4\Delta^2}{2} = \beta$ and transform to a new variate x defined by

$$x = \frac{C^2D^2}{n+C^2D^2}$$

Then x is distributed as

$$\text{Const.}\ (1-x)^{\frac{n-p-1}{2}} x^{\frac{p-2}{2}}\ {}_1F_1\left\{\frac{n+1}{2}, \frac{p}{2}, \beta x\right\} dx \ldots (4)$$

where the constant is easily found out to be

$$\text{Const} = \frac{\Gamma\frac{n+1}{2}}{\Gamma\frac{p}{2}\Gamma\frac{n-p+1}{2}} e^{-\beta} \quad (4\cdot5)$$

The distribution (4) involves three pramaters, n, p and Δ for different values of which we would have different tables of incomplete integral. There is a simple correspondence between the incomplete probability integral of D^2 and that of x, so that any one would serve the purpose of the other.

Putting now

$$I(n, p) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\frac{p}{2}\Gamma\left(\frac{n-p+1}{2}\right)} e^{-\beta}$$
$$\int_0^x (1-x)^{\frac{n-p-1}{2}} x^{\frac{(p-2)}{2}}\ {}_1F_1\left(\frac{n+1}{2}, \frac{p}{2}\beta x\right) dx \quad (5)$$

we have found that

$$I(n, p) = -\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\frac{p}{2}\Gamma\left(\frac{n-p+3}{2}\right)} e^{-\beta} (1-x)^{\frac{n-p+1}{2}}$$
$$x^{\frac{p-2}{2}}\ {}_1F_1\left(\frac{n+1}{2}, \frac{p}{2}, \beta x\right) + I(n, p-2) \quad (6)$$

Thus I (n,p) is seen to depend upon I (n,p − 2)

We come down therefore to two basic incomplete integrals I (n, 1), I (n, 2) according as p is odd or even. Now I (n, 2) has been found to be connected with I (n−2, 2) and I (n, 1) with I (n−2, 1) by the respective relations:

$$I(n,2) = e^{-\beta} x (1-x)^{\frac{n-3}{2}}\ {}_1F_1\left(\frac{n-1}{2}, 1, \beta x\right) + I(n-2,2) \quad (7)$$

$$I(n, 1) = \frac{\Gamma\left(\frac{n+1}{2}\right) e^{-\beta} (1-x)^{\frac{n-2}{2}} x^{\frac{1}{2}}}{\sqrt{\pi}\,\Gamma n/2}$$
$${}_1F_1\left(\frac{n-1}{2}, \tfrac{1}{2}\beta x\right) + I(n-2, 1) \quad (8)$$

The evalution of I (n, p) for any particular value of x depends therefore in the first place upon the use of tables for confluent hypergeometric function with which $_1F_1$ is simply connected and in the second place upon the evaluation by numerical quadrature of

I (2, 2); I (3, 2); I (1, 1); I (2, 1);

but luckily

$$I(3, 2) = x\ e^{\beta(x-1)}$$

and $$I(2, 1) = x^{\frac{1}{2}} e^{\beta(x-1)}$$

We are therefore left with the evaluation by numerical quadrature of the two basic incomplete integrals I (1, 1) and I (2, 2), whose forms were evident from the relation (5).

LETTERS TO THE EDITOR

The actual construction of numerical tables for (5) for different values of the variate x and the parameters n, p and β is now in progress at the Statistical Laboratory on the general mathematical plan indicated in this note. It should be noticed that by putting $R^2 = x$ in (4) we immediately get the well known distribution of the multiple correlation under a particular kind of hypothesis. Hence the proposed set of tables will also serve for the incomplete probability integral of this type of multiple correlation.

Statistical Laboratory, Calcutta, 6-5-1940.

S. N. Roy
Purnendu Bose

[1] Bose, R. C. and Roy, S. N. The distribution of the Studentised D^2 Statistic. *Proceedings of the Indian Statistical Conference*, Calcutta 1938.

[2] Roy, S. N. A Note on the distribution of the Studentised D^2-statistic. *Sankhya*. Vol. 4 Part 3.

[3] Roy, S. N. and Bose, R. C. The use and distribution of the Studentised D^2-statistic when variances and co-variances are based on K samples.—*Proceedings of the Second Indian Statistical Conference*, 1939.