Actes de la septieme
Rencontre Franco-Belge de Statisticiens
novembre 1986

THERE IS BETA AND BETA

Jacques Poitevineau et Bruno Lecoutre

GROUPE MATHEMATIQUES ET PSYCHOLOGIE
C.N.R.S. - U.A. 1201
12 RUE CUJAS, 75005 PARIS

THERE IS BETA AND BETA

Bruno LECOUTRE and Jacques POITEVINEAU
Groupe Mathématiques et Psychologie
C.N.R.S. - U.A. 1201
Université René Descartes
12 rue Cujas, 75005 Paris

1. A simple computation routine?

The Bayesian extensions of the analysis of variance involve several probability distributions that have not been extensively investigated at the present time (see Lecoutre, 1984). For practical applications we have undertaken to program these distributions on electronic computers. We will consider here the cumulative function of the "psi-square" distribution. The psi-square distribution extends both the classical Student's t and the non-central chi-square distributions (see Lecoutre, 1981; Rouanet and Lecoutre, 1983; Lecoutre, 1985) and can be characterized as follows:

Insert Definition

It might be thought that evaluating F(x) from an incomplete Beta function is a simple routine computation. Yet our investigations have quickly shown that there are serious difficulties in the use of the standard libraries currently available in C.I.R.C.E. (computation centre of the C.N.R.S. in Orsay).

2. First troubles

From the first results we noticed some serious abnormal functioning; even for the particular case $a^2 = 0$, where the probabilities F(x) are obtained by a single call to the incomplete Beta function. Of course, the problem got worse in the general case where multiple calls are needed. So we were led to compare the incomplete Beta routines of the available libraries (NAG, IBM-SSP, and IMSL). We then discovered puzzling results, which dramatically violate the well known properties of the Beta function. In particular it was not always the case that the probabilities, as expected, lay between 0 and 1 and were monotonous function of the parameters. As an illustration table 1 shows, with six decimal places, some probabilities returned by the routines G01BDE (NAG), BDTR (IBM-SSP), and MDBETA (IMSL). Computation time (in microseconds) is indicated in parenthesis . The parameters of the Beta distribution are p(=10) and q (varied); x denotes the value of the variable. Needless to say, the routine error codes indicated everything was all right, even in the worst cases (this remark applies throughout the text).

Table 1

Remark: since that time, G01BDE has been corrected (NAG Mark 11) and now successively returns: 0.066589, 0.535524, 1.000000, etc. But problems still remain (see later).

Apart from the differences between the routines, some abnormal values can be seen (even if some results are within the precision specified in the notice - four digits for G01BDE for instance). In some circumstances, the preceeding results could be sufficient; but, as in our case where those results were intermediate computations, the cumulated "errors" quickly lead to unusable final results.

On the other hand, we obtained satisfactory results with a routine called BETA based upon an algorithm by Majumder and Bhattacharjee (1973) (which is, surprinsingly, given as reference in the NAG documentation about G01BDE): cf Poitevineau, 1985.

The outputs of this routine for the preceeding examples are given in table 2.

Table 2
----

3. Getting worse

When the parameters are large (1) (i.e. exceed a limit chosen by the user) the first version BETA used the same approximations as those of BDTR: a chi-square approximation and, possibly, a subsequent normal approximation. Then, in cases where approximations stand, again bad results occurred. In a second version, we obtained approximations which looked good by using the method given by Peizer and Pratt (1968).

In table 3 are shown some important deviations exhibited by the approximations of BDTR.

Table 3

The first column (BETA) shows the results obtained without approximations which serve as a reference to which other results are compared (2); the third column (BDTR), with only one entry, permits to check the two routines agree when no approximations are used.

As can be seen, BETA approximations are very close (!) to the "reference" (we ensure we didn't run twice the routine without approximations; actually differences appear after the 7th digit), while those of BDTR are surprisingly different: 0.05, 0.11, 0.16 are hardly negligible deviations for probabilities!

For G01BDE (NAG Mark 11), we also noticed that the approximations seem to be less precise (especially when parameters are very different).

Examples are given in table 4 (here again the first column serve as the reference).

Table 4

For MDBETA the computation times let one think there is no approximation at all (results for the examples of tables 3 and 4 are close to those of BETA). But the results are not convincing when parameters become very large... (see table 1).

The BETA function subprogram (written in FORTRAN 77) has been implemented in the general library of C.I.R.C.E. (and also the routine PSI2D, PSI2, PSI2I, which respectively compute the psi-square probability density, the psi-square distribution function, and its inverse).

REFERENCES

Guigues, J.-L. (1982) - Note sur le calcul de la distribution du psi-deux. Paris: Groupe Mathématiques et Psychologie (unpublished).

Lecoutre, B. (1981) - Extensions de l'analyse de la variance: l'analyse bayésienne des comparaisons. Mathématiques et Sciences Humaines, 75, 49-69.

Lecoutre, B. (1984) - L'Analyse Bayésienne des Comparaisons. Lille: Presses Universitaires de Lille.

Lecoutre, B. (1985) - Reconsideration of the F test of the analysis of variance: The semi-bayesian significance test. Communications in Statistics, Theory and Methods, 14, 2437-2446.

Majumder, K.L., Battacharjee, G.P. (1973) - The incomplete Beta integral. Algorithm AS63. Applied Statistics, 22, 409-411.

Peizer, D.B., Pratt, J.W. (1968) - A normal approximation for Binomial, F, Beta and other common, related tail probabilities integral. Journal of the Americal Statistical Association, 63, 1416-1456.

Poitevineau, J. (1985) - Il y a bêta et bêta (ou NAG, IMSL, IBM et les autres). CIRCE Interface, 67, 2341-2342.

Rouanet, H., Lecoutre, B. (1983) - Specific inference in ANOVA: From significance tests to Bayesian procedures. British Journal of Mathematical and Statistical Psychology, 36, 252-268.

FOOTNOTES

(1) Even if large values can seem to be rare in practice, they may occur during the computation of the psi-square distribution function, since the algorithm involves successive Beta distributions with increasing parameters. they also can be found in the case of Bayesian inference on frequencies.

(2) A simple normal approximation gives .8413 (for all rows of course).

If u is distributed $t_{p;q}(a, I_p)$ (p-variate t distribution with $q \leqq p$ degrees of freedom, centred around **a** and of unit scale matrix I_p) then the variable $\psi^2 = u'u$ has a psi-square distribution with p and q degrees of freedom and with eccentricity $a^2 = \mathbf{a}'\mathbf{a}$, which we denote $\psi^2 \sim \psi^2_{p,q}(a^2)$.

Hence the immediate properties (symbolically written):

$\psi^2_{p,\infty}(a^2) = \chi^2_p(a^2)$ (non-central chi-square distribution)

$\psi^2_{p,q}(0) = pF_{p,q}$ (central F distribution)

From the density function $p(\psi^2)$ of the psi-square distribution, we can write its cumulative distribution function as a mixture of usual incomplete Beta integrals (see Guigues, 1982):

$$F(x) = \int p(\psi^2)d\psi^2 = \sum_{j=0}^{+\infty} c_j I_y(j+\frac{p}{2}, j+\frac{q}{2})$$

$$\text{with } c_j = \frac{1}{j!}\frac{\Gamma(j+\frac{q}{2})}{\Gamma(\frac{q}{2})}\left(\frac{q}{q+a^2}\right)^{\frac{q}{2}}\left(\frac{a^2}{q+a^2}\right)^j$$

$$y = \frac{x}{q+a^2+x}$$

where $I_y(j+\frac{p}{2}, j+\frac{q}{2})$ is the usual incomplete Beta integral

DEFINITION

p=10	G01BDE (NAG)	BDTR (IBM-SSP)	MDBETA (IMSL)
x=.1 q=50	0.066589 (321)	0.066589 (1042)	0.066589 (602)
x=.1 q=90	0.535605 (349)	0.535524 (986)	0.535523 (727)
x=.3 q=90	1.000144 (337)	1.000000 (986)	0.999999 (611)
x=.3 q=100	1.000009 (343)	1.000000 (984)	0.999999 (643)
x=.3 q=1010	1.000000 (195)	1.000000 (634)	0.999963 (3436)
x=.3 q=2000	1.000000 (192)	1.000000 (626)	0.999888 (6301)
x=.3 q=10000	1.000000 (193)	1.000000 (322)	0.999292 (29404)

TABLE 1

p=10	BETA
x=.1 q=50	0.066589 (287)
x=.1 q=90	0.535524 (292)
x=.3 q=90	0.999999 (292)
x=.3 q=100	1.000000 (274)
x=.3 q=1010	1.000000 (384)
x=.3 q=2000	1.000000 (421)
x=.3 q=10000	1.000000 (270)

TABLE 2

x=mean+standard deviation	BETA	BETA*	BDTR	BDTR*
x=.345501 p=500 q=1000	0.841277	--	0.841277	--
x=.344363 p=500 q=1005	0.841276	0.841276	--	0.892212
x=.343048 p=1005 q=2000	0.841309	0.841309	--	0.955405
x=.094068 p=1005 q=10000	--	0.841353	--	0.999411

* with use of approximation

TABLE 3

x=mean+standard deviation	BETA	BETA*	G01BDE*
x=.001200 p=.5 q=1005	0.879661	0.879654	0.884005
x=.003388 p=2 q=1005	0.854605	0.854605	0.856084
x=.344363 p=500 q=1005	0.841276	0.841276	0.841281

* with use of approximation

TABLE 4