Algorithm AS 200

Approximating the Sum of Squares of Normal Scores

By N. Balakrishnan

Indian Institute of Technology, New Delhi, India

[Received November 1980; Final revision February 1983]

Keywords: Order statistics; Normal scores; Legendre polynomial; Fourier coefficient; Approximation and bounds

Language

Fortran 66

Description and Purpose

Following the notation of Royston (1982), let $E(i, N)$ denote the mean of the ith order statistic obtained from a random sample of size N from a standard normal distribution. An algorithm is presented to approximate the sum,

$$S_N = \sum_{i=1}^{N} \{E(i, N)\}^2,$$

for a given value of N and also to give bounds for the truncation error involved in the approximation procedure.

The sum of squares of normal scores is of great importance in the analysis of ranked data. In addition, the algorithm given recently by Davis and Stephens (1978) for approximating the variance-covariance matrix of normal order statistics also requires a knowledge of S_N. Approximations for S_N have been proposed and discussed in great detail by Ruben (1956) and Saw and Chow (1966). A simple method, noted in Joshi and Balakrishnan (1983), is used to obtain an approximation and bounds for S_N. The Fourier coefficients involved in the procedure have been taken from Table 1 of Joshi and Balakrishnan (1983).

Numerical Method

Let $L_k(u)$ be the kth orthonormal Legendre polynomial in $[0, 1]$ given by

$$L_k(u) = (2k+1)^{1/2} (k!)^{-1} \frac{d^k}{du^k} \{u^k(u-1)^k\}, \quad k = 0, 1, 2, \ldots.$$

With $u = \Phi(x)$ as the cdf of a standard normal variate, let a_k be the Fourier coefficient of $\Phi^{-1}(u)$ with respect to $L_k(u)$. Then, it is well known that

$$\sum_{i=1}^{N} \{E(i, N)\}^2 = \sum_{k=0}^{N-1} a_k^2 \frac{(N!)^2}{(N+k)!\,(N-k-1)!}$$

Also for the standard normal distribution, we have $a_{2k} = 0$ $(k = 0, 1, 2, \ldots)$, see, for example, Joshi and Balakrishnan (1983). Therefore, upon writing

$$\sum_{i=1}^{N} \{E(i, N)\}^2 = \sum_{k=0}^{2m-1} a_k^2 \frac{(N!)^2}{(N+k)!\,(N-k-1)!} + E_{2m-1}^{(N)},$$

Present address: Department of Statistics, The University of Manitoba, Winnipeg, Manitoba, Canada.

 0035–9254/84/33242 $2.00

where $E^{(N)}_{2m-1}$ stands for the error due to truncation, it can easily be seen that (see also Saw and Chow, 1966)

$$F_{N,m} \quad a^2_{2m+1} \leqslant E^{(N)}_{2m-1} \leqslant F_{N,m} \left[1 - \sum_{j=0}^{2m-1} a_j^2 \right],$$

$$F_{N,m} = N \frac{(N-1)(N-2)\ldots(N-2m-1)}{(N+1)(N+2)\ldots(N+2m+1)}.$$

Finally, writing

$$L_k(u) = \sum_{i=0}^{k} (-1)^{k-i} (2k+1)^{1/2} \binom{k}{i}^2 u^i (1-u)^{k-i}$$

and using the symmetry about the origin of the standard normal distribution, we have for $k = 0, 1, 2, \ldots$,

$$a_{2k+1} = (4k+3)^{1/2} (k+1)^{-1} \sum_{i=0}^{k} (-1)^{i+1} \binom{2k+1}{i} E(i+1, 2k+2).$$

Using the ten-figure tables of the moments of normal order statistics by Yamauti (1972), these coefficients have been tabulated for $k = 0\ (1)\ 13$ and given in Table 1 of Joshi and Balakrishnan (1983).

Structure

SUBROUTINE SUMSQ(N, APPROX, BL, BU, IFAULT)

Formal parameters

N	Integer	input:	size of the sample
APPROX	Real	output:	approximate value of the sum of squares of normal scores in a sample of size *N*
BL	Real	output:	lower bound for the error due to truncation in the approximation
BU	Real	output:	upper bound for the error due to truncation in the approximation
IFAULT	Integer	output:	a fault indicator, equal to 1 if $N \leqslant 1$ 0 otherwise

Accuracy and Time

As can be seen from the method described, the results for $N \leqslant 27$ will be exact, except for possible errors in the computation. The truncation error is found to be at most 2×10^{-9} for $N \leqslant 50$, 9×10^{-6} for $50 < N \leqslant 100$, 9×10^{-4} for $100 < N \leqslant 200$ and 4.5×10^{-3} for $200 < N \leqslant 300$. However, for higher values of N, the error tends to be of larger magnitude as one would expect. Thus, for example, for $N = 1000$, *BL* and *BU* are obtained respectively as 0.0164 and 0.0893, and so the approximate value that is obtained may be accurate only up to the first decimal place. If higher accuracy is required in this range of N, it may therefore be necessary to include some more terms in the approximate series and use a suitably modified version of *SUMSQ*.

The following table gives the approximations obtained by the routine and also the tabulated values for some selected choices of N. This is to get an idea about the efficiency of the proposed method. For $N = 5$, 10, 20 and 27, the tabulated values are given after calculating them directly

TABLE 1

N	*Approximation*	*Tabulated value*	*Time (seconds)*
5	3.19506030	3.19506030	0.03
10	7.91427186	7.91427186	0.04
20	17.67818072	17.67818073	0.04
27	24.58794948	24.58794949	0.04
50	47.42169559	47.42170	0.04
75	72.32423601	72.32427	0.04
100	97.25999311	97.25999	0.04

from the table of means of normal order statistics provided by Yamauti (1972). For $N = 50$, 75 and 100, the values are taken from the table of squares of normal scores given in Pearson and Hartley (1972, p. 217).

Time, as expected, does not depend very much on N. However, for the sake of completeness, the time required for the computation on a DEC 1090 computer is also provided in the table for each choice of N.

Acknowledgements

The author wishes to thank the referees for their valuable comments and suggestions in the preparation of this paper.

References

Davis, C. S. and Stephens, M. A. (1978) Algorithm AS 128. Approximating the covariance matrix of normal order statistics. *Appl. Statist.*, **27**, 206–212.
Joshi, P. C. and Balakrishnan, N. (1983) Bounds for the moments of extreme order statistics for large samples. *Mathematische Operationsforschung und Statistik, Series Statistics*, **14**, 387–396.
Pearson, E. S. and Hartley, H. O. (ed.) (1972) *Biometrika Tables for Statisticians*, Vol. II. London: Cambridge University Press.
Royston, J. P. (1982) Algorithm AS 177. Expected normal order statistics (exact and approximate). *Appl. Statist.*, **31**, 161–165.
Ruben, H. (1956) On the sum of squares of normal scores. *Biometrika*, **43**, 456–458. Correction, **52**, 669.
Saw, J. G. and Chow, B. (1966) The curve through the expected values of ordered variates and the sum of squares of normal scores. *Biometrika*, **53**, 252–255.
Yamauti, Z. (ed.) (1972) *Statistical Tables and Formulas with Computer Applications*, JSA-1972. Tokyo: Japanese Standard Association.

```
      SUBROUTINE SUMSQ(N, APPROX, BL, BU, IFAULT)
C
C        ALGORITHM AS 200  APPL. STATIST. (1984) VOL.33, NO.2
C
C        APPROXIMATES THE SUM OF SQUARES OF NORMAL SCORES AND
C        GIVES BOUNDS FOR THE TRUNCATION ERROR INVOLVED IN
C        THE APPROXIMATION
C
C        COEFFICIENTS SA(K) ARE THE SQUARES OF THE FOURIER
C        COEFFICIENTS A(2*K-1)
C
      REAL APPROX, BL, BU, SA(14), ZERO, ONE, PROD, RI, RK, RN, SUM1,
     *  SUM2
C
      DATA SA(1), SA(2), SA(3), SA(4), SA(5), SA(6), SA(7), SA(8),
     *  SA(9), SA(10), SA(11), SA(12), SA(13), SA(14) /0.9549296583,
     *  0.0334920160, 0.0066747227, 0.0022780925, 0.0010163627,
     *  0.0005326443, 0.0003108437, 0.0001958978, 0.0001308187,
     *  0.0000913941, 0.0000663703, 0.0000496799, 0.0000411471,
     *  0.0000349349/, ZERO /0.0/, ONE /1.0/, NMAX /27/
C
C        VALUE OF ZERO IS ARBITRARILY GIVEN IN CASE OF A FAILURE
C
      APPROX = ZERO
      BL = ZERO
      BU = ZERO
```

```
C         CHECK CONSISTENCY OF INPUT PARAMETER
C
      IFAULT = 1
      IF (N .LE. 1) RETURN
      IFAULT = 0
      RN = N
      NM = NMAX - 2
      IF (N .GT. NMAX) GOTO 1
      M = N / 2
      NM = 2 * M - 1
    1 SUM1 = ZERO
      SUM2 = ZERO
      KK = 0
      DO 3 K = 1, NM, 2
      RK = K
      PROD = ONE
      DO 2 I = 1, K
      RI = I
      PROD = PROD * (RN - RI) / (RN + RK - RI + ONE)
    2 CONTINUE
      KK = KK + 1
      PROD = PROD * RN
      SUM1 = SUM1 + SA(KK) * PROD
      SUM2 = SUM2 + SA(KK)
    3 CONTINUE
C
C         APPROXIMATE VALUE IS COMPUTED
C
      APPROX = SUM1
      IF (N .LE. NMAX) RETURN
      PROD = ONE
      DO 4 I = 1, NMAX
      RI = I
      PROD = PROD * (RN - RI) / (RN + RI)
    4 CONTINUE
      PROD = PROD * RN
C
C         LOWER AND UPPER BOUNDS FOR THE TRUNCATION ERROR ARE COMPUTED
C
      BL = PROD * SA(14)
      BU = PROD * (ONE - SUM2)
      RETURN
      END
```

Algorithm AS 201

Combined Significance Test of Differences Between Conditions and Ordinal Predictions

By Ranald R. Macdonald

University of Stirling, Scotland

[Received February 1983; Revised July 1983]

Keywords: Significance test; Combined tests; Ordinal predictions

Language

Fortran 66

Description and Purpose

This paper presents an algorithm for incorporating an experimenter's beliefs about what sorts

Present address:

 0035–9254/84/33245 $2.00