```
C
C        CALCULATION BY CONTINUED FRACTION
C
   30 A=1.0-P
      B=A+X+1.0
      TERM=0.0
      PN(1)=1.0
      PN(2)=X
      PN(3)=X+1.0
      PN(4)=X*B
      GIN=PN(3)/PN(4)
   32 A=A+1.0
      B=B+2.0
      TERM=TERM+1.0
      AN=A*TERM
      DO 33 I=1,2
   33 PN(I+4)=B*PN(I+2)-AN*PN(I)
      IF(PN(6).EQ.0.0) GO TO 35
      RN=PN(5)/PN(6)
      DIF=ABS(GIN-RN)
      IF(DIF.GT.ACU) GO TO 34
      IF(DIF.LE.ACU*RN) GO TO 42
   34 GIN=RN
   35 DO 36 I=1,4
   36 PN(I)=PN(I+2)
      IF(ABS(PN(5)).LT.OFLO) GO TO 32
      DO 41 I=1,4
   41 PN(I)=PN(I)/OFLO
      GO TO 32
   42 GIN=1.0-FACTOR*GIN
C
   50 GAMAIN=GIN
      RETURN
      END
```

Algorithm AS 33

Calculation of Hypergeometric Sample Sizes

By F. B. Leech

Rothamsted Experimental Station, Harpenden, Herts

Language

ASA Standard Fortran

Purpose

Solution, in n, of $P = \{(N-m)!\,(N-n)!\}/\{N!\,(N-m-n)!\}$ for given values of P, N and $p = m/N$.

The program was originally written in connection with a paper on innocuity testing of inactivated foot-and-mouth disease vaccines (Anderson *et al.*, 1970). In this context, a sample of n cm^3 from a vaccine batch of N cm^3 is tested for the presence of residual active virus. Batches giving any reaction are rejected and the theoretical problem is to select a size of sample such that when the test is negative, the amount, if any, of active virus in the batch must be small. In this context, the theory involves only the first term of the hypergeometric distribution. In the equation above, P is the expected proportion of samples of n cm^3 negative to the test, when the batch of N cm^3 contains m infective cm^3. The right-hand side is the first term of the hypergeometric distribution; corresponding equations, when $1, 2, \ldots$ cm^3 react to the test, involve the second and later terms of the distribution and the problem of solving them would be much more difficult.

The method uses the fact that $P = \{1-(n/N)\}^m$ is a very close approximation to the hypergeometric equation. Starting with trial values $n = 0$ and NP, a new value of n is obtained by interpolation on the line $\log(N-n) = a + b \log P$. This process is repeated using the two most recent values of n until two successive estimates are the same. Within the range of parameters tested, the first interpolation was usually correct and never differed by more than one unit from the correct solution, which was reached on the next iteration. For many purposes, the simpler formula $P = \{1-(n/N)\}^m$, which can readily be solved without a computer program, has all necessary accuracy.

Structure

SUBROUTINE HYPERG (N,P,PP,K,IFAULT)

Formal parameters

N	Integer	input:	as in the formula above.
P	Real	input:	the probability P.
PP	Real	input:	the proportion m/N.
K	Integer	output:	the solution.
IFAULT	Integer	output:	fault indicator.

Subroutine *HYPERG* has a function *FLOG* to calculation $\log X!$, using Stirling's approximation $\frac{1}{2}\log(2\pi) + (X+\frac{1}{2})\log X - X + 1/(12X)$.

Failure indication

IFAULT = 1 if after 9 iterations the solution has not been reached;

= 2 if a proportion supplied by the external program not between 0 and 1;

= 0 otherwise.

Restrictions

None.

Accuracy

Results for small m can be checked arithmetically. These were found to be the nearest integral values to the non-integral exact solution.

Reference

Anderson, E. C., Capstick, P. B., Mowat, G. N. and Leech, F. B. (1970). A comparison between the use of cattle and a tissue culture system for safety testing of inactivated foot-and-mouth disease vaccines. *J. Hyg., Camb.*, **68**, 159–172.

```
      SUBROUTINE HYPERG (N,PP,P,K,IFAULT)
C
C        ALGORITHM AS 33 J.R.STATIST.SOC. C. (1970) VOL.19 NO.3
C
C        GIVEN N = NUMBER OF UNITS OF BATCH TO BE SAMPLED PP =
C        PROBABILITY OF POSITIVE UNITS BEING PRESENT, AND P= PROBABILITY
C        OF FAILING TO DETECT POSITIVE UNITS, FIND K, THE SIZE OF
C        SAMPLE TO BE DRAWN FROM THE BATCH
C
      DOUBLE PRECISION A(10),B(10),BB,C,D,DBLE,X,FLOG,DLOG,DEXP
C
      IFAULT=0
      IF((P.GT.1.0).OR.(P.LT.0.0)) GO TO 4
      IF((PP.GT.1.0).OR.(PP.LT.0.0)) GO TO 4
      X=N
      D=DLOG(DBLE(P))
      M=FLOAT(N)*PP+0.5
      A(1)=0.D0
      B(1)=DLOG(X)
      IF((P+PP).GT.1.) P=1.-PP
    2 K=FLOAT(N)*P+0.5
      I=N-K
      X=I
      C=DLOG(X)
      N1=N-M
      DO 3 L=2,10
      N3=N1-K
      A(L)=FLOG(N1)+FLOG(I)-FLOG(N)-FLOG((N3))
      B(L)=C
      BB=(B(L)-B(L-1))/(A(L)-A(L-1))
      C=B(L)+BB*(D-A(L))
      I=DEXP(C)+0.5D0
      K1=N-I
      IF(K.EQ.K1) RETURN
      K=K1
    3 CONTINUE
      IFAULT=1
C     SET DIAGNOSTIC FOR NON-CONVERGENCE
      RETURN
    4 IFAULT=2
C        SET DIAGNOSTIC FOR DATA ERROR (PROPORTION NOT BETWEEN 0 AND 1)
      RETURN
      END
C
      DOUBLE PRECISION FUNCTION FLOG(N)
C        GIVEN N, FIND LOG FACTORIAL N USING STIRLING'S APPROXIMATION
      DOUBLE PRECISION F(6),X,DLOG
      DATA F(1),F(2),F(3),F(4),F(5),F(6)/0.D0,0.D0,.6931478D0,
     1 1.7917595D0,3.1780538D0,4.7874917D0/
      IF(N.GT.5) GO TO 1
      FLOG=F(N+1)
      RETURN
    1 X=N
      FLOG=.91893854D0+(X+.5D0)*DLOG(X)-X+1.0D0/(12.D0*X)
      RETURN
      END
```