National Bureau of Standards
Statistical Engineering Division
Gaithersburg, MD 20899

N-86-3
October 1986

An Algorithm for Computing the Beta C.D.F. to a Specified Accuracy

Charles P. Reeve

The cumulative distribution function (c.d.f.) of the beta distribution (also called the incomplete beta ratio) is

$$I(x,p,q) = \int_0^x t^{p-1}(1-t)^{q-1}dt/B(p,q) \quad (0<x<1;p>0;q>0) \qquad (1)$$

where $B(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1}dt$. [The left hand side of (1) is not standard statistical notation but is used here for convenience.] This function is very important in statistical computing because the c.d.f.'s of the t, noncentral t, F, noncentral F, binomial, and negative binomial distributions are easily obtainable from it. For example,

$$t \rightarrow P\{t(\nu)<x\} = \{1 + \text{sign}(x)I(x^2/[x^2+\nu],1/2,\nu/2)\}/2$$

$$F \rightarrow P\{F(\nu_1,\nu_2)<x\} = I(\nu_1 x/[\nu_1 x+\nu_2],\nu_1/2,\nu_2/2),$$

$$Bin \rightarrow P\{Bi(n,p)<x\} = \begin{cases} I(1-p,n-x,x+1) & (x=0,1,2,\ldots,n-1) \\ 1 & (x=n) \end{cases} \text{, and}$$

$$NB \rightarrow P\{NBi(x,p)<y\} = I(p,x,y+1) \quad (x=1,2,3,\ldots;y=0,1,2,\ldots).$$

Recent papers by Bosten and Battiste [2], Koo[4], and Majumber and Bhattacharjee [5] have served as a basis for computer algorithms currently available at NBS [3,6,7]. Half-integer arguments are required in [4], and the implementation of [2] in CMLIB [6] was found to give erroneous results for certain large values of p and q.

The algorithm described in this note is a modification of the algorithm in [5] which, in the author's opinion, has two shortcomings:

1) The criterion for truncating the infinite series is based on the first omitted term rather the sum of all the omitted terms, and

2) When x is near p/(p+q), for certain p and q, a very large number of terms must be summed before achieving the required accuracy.

The modified algorithm, which has been strengthened in these two areas, is based on the two recurrence relations

$$I(x,p,q) = \frac{x^p(1-x)^{q-1}}{pB(p,q)} + I(x,p+1,q-1) \quad (p>0;q>1) \tag{2}$$

$$I(x,p,q) = \frac{x^p(1-x)^q}{pB(p,q)} + I(x,p+1,q) \quad (p>0;q>0) \tag{3}$$

given by equations 26.5.15 and 26.5.16 of Abramowitz and Stegun[1]. In both cases the lower tail area of one beta distribution is expressed in terms of the lower tail area of another beta distribution. Applying (2) n times yields

$$I(x,p,q) = \sum_{i=1}^{n} \frac{x^{p+i-1}(1-x)^{q-i}}{(p+i-1)B(p+i-1,q-i+1)} + I(x,p+n,q-n) \tag{4}$$

provided that $0<n<q$. Applying (3) m times to the rightmost term in (4),

$$I(x,p,q) = \sum_{i=1}^{n} \frac{x^{p+i-1}(1-x)^{q-i}}{(p+i-1)B(p+i-1,q-i+1)} + \sum_{j=1}^{m} \frac{x^{p+n+j-1}(1-x)^{q-n}}{(p+n+j-1)B(p+n+j-1,q-n)} + I(x,p+n+m,q-n) \tag{5}$$

where $m>0$. For a given $x<1$, p, q, and absolute accuracy limit $\varepsilon>0$ there will be non-negative integers $n<q$ and m such that $I(x,p+n+m,q-n)<\varepsilon$ (the proof is left to the reader). The summation of n+m terms will then be an acceptable approximation to $I(x,p,q)$. Note that the second summation in (5) is needed only if the first summation is carried to completion $(0<q-n<1)$ without achieving the desired absolute accuracy.

An upper bound for $I(x,p,q)$ is computed by applying (3) repeatedly to obtain

$$I(x,p,q) = \frac{x^p(1-x)^q}{pB(p,q)}\left[1 + \sum_{i=1}^{\infty} \frac{(p+q)(p+q+1)\ \ldots\ (p+q+i-1)}{(p+1)(p+2)\ \ldots\ (p+i)} x^i\right]. \tag{6}$$

If $q<1$ then from (6)

$$\begin{aligned} I(x,p,q) &< \frac{x^p(1-x)^q}{pB(p,q)}\left[1 + \sum_{i=1}^{\infty} x^i\right] \\ &< \frac{x^p(1-x)^q}{pB(p,q)(1-x)}, \end{aligned} \tag{7}$$

when $x<1$. If $q>1$ then

$$\begin{aligned} I(x,p,q) &< \frac{x^p(1-x)^q}{pB(p,q)}\left[1 + \sum_{i=1}^{\infty} \left(\frac{p+q}{p+1}\right)^i x^i\right] \\ &< \frac{x^p(1-x)^q}{pB(p,q)[1-\ (p+q)x/(p+1)]} \end{aligned} \tag{8}$$

when $x<(p+1)/(p+q)$. Combining the bounds in (7) and (8) yields the bound for the remainder term in (5),

$$I(x,p+n+m,q-n) < \frac{x^{p+n+m}(1-x)^{q-n}}{(p+n+m)B(p+n+m,q-n)(1-x/r)} \tag{9}$$

where $r=\min[1,(p+n+m+1)/(p+q+m)]$ provided that $x<r$. When the right hand side of (9) becomes $<\varepsilon$ as n or m increases, the series which is being summed is truncated.

Although the above algorithm works for $0<x<1$, the number of terms required may become very large as $x\to1$. In this case a more efficient procedure is to compute $I(1-x,q,p)$ and then use the identity

$$I(x,p,q) = 1 - I(1-x,q,p).$$

This is done in [5] when $x>p/(p+q)$. However, when x is slightly less than this cutoff value, a large number of terms may be required for convergence. The author conducted a study using cutoff values of the form $(p+\omega)/(p+q+2\omega)$ and found that the value $\omega=20$ did the best job of minimizing the number of terms required for convergence for a broad range of p and q values. The increased efficiency using this new cutoff value is illustrated in table 1 for the case p=10 and q=1.

The above algorithm with $\omega=20$ has been implemented in the FORTRAN subroutine CDFBET. If $x<0$ then 0 is returned, and if $x>1$ then 1 is returned. If x is in the extreme tails of the distribution (near 0 or 1), exponential underflow might occur when computing the first term (i=1) in (5). This is prevented by testing the exponent to see if it is less than a pre-set underflow limit. If it is then $I(x,p,q)$ is set to 0 or 1 as appropriate. An external routine, such as Reeve[8] must be provided for computing the log of the gamma function in double precision.

A listing of CDFBET is an appendix to this note. It is invoked by

```
CALL CDFBET(X,P,Q,EPS,IFLAG,CDFX)
```

where the variable names are defined in the program documentation. The returned value of CDFX is valid only if IFLAG=0 on return. In passing ε (variable name EPS) to CDFBET the user should realize that accuracy is limited by the number of digits carried in a single precision variable, and that roundoff error may affect the last one or two of these digits.

Table 1

Comparison of the number of terms (n+m) required for convergence of the beta c.d.f. using two different cutoff criteria (ω). The integers n and m are as in (4) and (5).

$p=10 \quad q=1 \quad \varepsilon=10^{-4}$

cutoff value	x	c.d.f.(x)	$\omega=0$			$\omega=20$		
			n	m	n+m	n	m	n+m
	0.10	0.0000	0	0	0	0	0	0
	0.20	0.0000	0	0	0	0	0	0
	0.30	0.0000	0	0	0	0	0	0
	0.40	0.0001	0	1	1	0	1	1
	0.50	0.0009	0	4	4	0	4	4
0.5882 →→→						----	----	----
(ω=20)	0.60	0.0060	0	9	9	9	1	10
	0.70	0.0282	0	16	16	9	0	9
	0.80	0.1073	0	32	32	7	0	7
	0.90	0.3486	0	78	78	6	0	6
0.9091 →→→			----	----	----			
(ω=0)	0.95	0.5988	4	0	4	4	0	4

$p=10 \quad q=1 \quad \varepsilon=10^{-8}$

cutoff value	x	c.d.f.(x)	$\omega=0$			$\omega=20$		
			n	m	n+m	n	m	n+m
	0.10	0.00000000	0	0	0	0	0	0
	0.20	0.00000010	0	2	2	0	2	2
	0.30	0.00000590	0	6	6	0	6	6
	0.40	0.00010485	0	11	11	0	11	11
	0.50	0.00097656	0	17	17	0	17	17
0.5882 →→→						----	----	----
(ω=20)	0.60	0.00604662	0	27	27	9	11	20
	0.70	0.02824753	0	42	42	9	6	15
	0.80	0.10737418	0	73	73	9	2	11
	0.90	0.34867844	0	165	165	9	0	9
0.9091 →→→			----	----	----			
(ω=0)	0.95	0.59873694	7	0	7	7	0	7

References

1. Abramowitz, Milton and Stegun, Irene A., *Handbook of Mathematical Functions*, NBS Applied Mathematics Series 55, 1970, p. 944.

2. Bosten, Nancy E. and Battiste, E.L., "Remark on Algorithm 179", *Communications of the ACM*, Vol. 17, No. 3, 1974, pp. 156-7.

3. IMSL, Inc., Houston, TX. [MDBETA]

4. Koo, Joo O., "Algorithm to Evaluate Incomplete Beta Function Ratio", *1984 Proceedings of the Statistical Computing Section*, American Statistical Association, p. 195-7.

5. Majumber, K.L. and Bhattacharjee, G.P., "The Incomplete Beta Integral", Algorithm AS63, *Applied Statistics*, Vol. 22, No. 3, 1973, pp. 409-11.

6. NBS Core Math Library (CMLIB). [BETAI]

7. Numerical Algorithms Group (NAG), Downers Grove, IL. [G01BDE]

8. Reeve, Charles P., "Accurate Computation of the Log of the Gamma Function", SED Note 86-1, October 1986.

```
      SUBROUTINE CDFBET (X,P,Q,EPS,IFLAG,CDFX)
C
C-----------------------------------------------------------------------
C   CDFBET   WRITTEN BY CHARLES P. REEVE, STATISTICAL ENGINEERING
C            DIVISION, NATIONAL BUREAU OF STANDARDS, GAITHERSBURG,
C            MARYLAND  20899
C
C   FOR: COMPUTING THE CUMULATIVE DISTRIBUTION FUNCTION OF THE BETA
C        DISTRIBUTION (ALSO KNOWN AS THE INCOMPLETE BETA RATIO) TO A
C        SPECIFIED ACCURACY (TRUNCATION ERROR IN THE INFINITE SERIES).
C        THE ALGORITHM, DESCRIBED IN REFERENCE 2, IS A MODIFICATION OF
C        THE ALGORITHM OF REFERENCE 1.  THREE FEATURES HAVE BEEN ADDED:
C
C        1) A PRECISE METHOD OF MEETING THE TRUNCATION ACCURACY,
C        2) A CONSTANT W USED IN DETERMINING FOR WHICH X VALUES THE
C           RELATION I(X,P,Q) = 1 - I(1-X,Q,P) IS TO BE USED, AND
C        3) A CONSTANT UFLO >= THE UNDERFLOW LIMIT ON THE COMPUTER.
C
C   SUBPROGRAMS CALLED: GAMLOG (LOG OF GAMMA FUNCTION)
C
C   CURRENT VERSION COMPLETED OCTOBER 24, 1986
C
C   REFERENCES:
C
C   1) MAJUMDER, K.L. AND BHATTACHARJEE, G.P., 'THE INCOMPLETE BETA
C      INTEGRAL', ALGORITHM AS 63, APPLIED STATISTICS, VOL. 22, NO. 3,
C      1973, PP. 409-411.
C
C   2) REEVE, CHARLES P., 'AN ALGORITHM FOR COMPUTING THE BETA C.D.F.
C      TO A SPECIFIED ACCURACY', STATISTICAL ENGINEERING DIVISION
C      NOTE 86-3, OCTOBER 1986.
C-----------------------------------------------------------------------
C   DEFINITION OF PASSED PARAMETERS:
C
C      * X = VALUE AT WHICH THE C.D.F. IS TO BE COMPUTED (REAL)
C
C      * P = FIRST PARAMETER OF THE BETA FUNCTION (>0) (REAL)
C
C      * Q = SECOND PARAMETER OF THE BETA FUNCTION (>0) (REAL)
C
C    * EPS =  THE DESIRED ABSOLUTE ACCURACY OF THE C.D.F. (>0) (REAL)
C
C    IFLAG = ERROR INDICATOR ON OUTPUT (INTEGER)   INTERPRETATION:
C            0 -> NO ERRORS DETECTED
C            1 -> EITHER P OR Q OR EPS IS <= UFLO
C            2 -> NUMBER OF TERMS EVALUATED IN THE INFINITE SERIES
C                 EXCEEDS JMAX
C
C     CDFX = THE C.D.F. EVALUATED AT X (REAL)
C
C   * INDICATES PARAMETERS REQUIRING INPUT VALUES
C-----------------------------------------------------------------------
C
```

```
      LOGICAL LL
      DOUBLE PRECISION DP,DQ,GAMLOG
      DATA IMAX,W,UFLO / 5000,20.0,1E-100 /
      CDFX = 0.0
C
C---- CHECK FOR VALIDITY OF ARGUMENTS P, Q, AND EPS
C
      IF (P.LE.UFLO.OR.Q.LE.UFLO.OR.EPS.LE.UFLO) THEN
         IFLAG = 1
         RETURN
      ENDIF
      IFLAG = 0
C
C---- CHECK FOR SPECIAL CASES OF X
C
      IF (X.LE.0.0) RETURN
      IF (X.GE.1.0) THEN
         CDFX = 1.0
      ELSE
C
C---- SWITCH ARGUMENTS IF NECESSARY
C
         LL = P+W.GE.(P+Q+2.0*W)*X
         IF (LL) THEN
            XY = X
            YX = 1.0-XY
            PQ = P
            QP = Q
         ELSE
            YX = X
            XY = 1.0-YX
            QP = P
            PQ = Q
         ENDIF
C
C---- EVALUATE THE BETA P.D.F. AND CHECK FOR UNDERFLOW
C
         DP = DBLE(PQ-1.0)*DLOG(DBLE(XY))-GAMLOG(PQ)
         DQ = DBLE(QP-1.0)*DLOG(DBLE(YX))-GAMLOG(QP)
         PDFL = SNGL(GAMLOG(PQ+QP)+DP+DQ)
         IF (PDFL.LT.ALOG(UFLO)) THEN
         ELSE
            U = EXP(PDFL)*XY/PQ
            R = XY/YX
   10       IF (QP.LE.1.0) GO TO 20
C
C---- INCREMENT PQ AND DECREMENT QP
C
            IF (U.LE.EPS*(1.0-(PQ+QP)*XY/(PQ+1.0))) GO TO 40
            CDFX = CDFX+U
            PQ = PQ+1.0
            QP = QP-1.0
            U = QP*R*U/PQ
```

betapdf = $$\frac{x^{(A-1)} \cdot (1-x)^{(B-1)}}{B(A,B)}$$

```
              GO TO 10
   20         V = YX*U
              YXEPS = YX*EPS
C
C--- INCREMENT PQ
C
              DO 30 J = 0, JMAX
                 IF (V.LE.YXEPS) GO TO 40
                 CDFX = CDFX+V
                 PQ = PQ+1.0
                 V = (PQ+QP-1.0)*XY*V/PQ
   30         CONTINUE
              IFLAG = 2
           ENDIF
   40      IF (.NOT.LL) CDFX = 1.0-CDFX
        ENDIF
        RETURN
        END
```