IIB: Inverse Incomplete Beta Function of One Half

by Jon F. Claerbout

Given N independent samples X_i , $i = 1,N$ from an unknown probability function, let $X_{(i)}$ denote the samples reordered from smallest to largest. Then if N is odd, $X_{((i+1)/2)}$ is an unbiased estimate of the median in the sense that regardless of the underlying probability function, it turns out that the probability is one half that $X_{((i+1)/2)}$ underestimates the true unknown median and half that it overestimates it. The inverse incomplete beta function of one half provides the answer to the more general question, what quantile does $X_{(i)}$ estimate? This property of an estimate, that it is equally likely to be greater or less than the true value is called unbiased in the median. More descriptive background is in "Probability and Entropy of seismic Data" (this report, p. 101).

The Handbook of Mathematical Functions by Abramovitz and Stegun defines the binomial distribution in terms of the incomplete beta function on page 945 in equation 26.5.24 as

$$\sum_{s=a}^{n} \binom{n}{s} p^s (1-p)^{n-s} = I_p(a, n-a+1) \qquad (1)$$

What we want to do is to set this equal to one half and solve for $p(a,n)$. Then we identify n with N and a with i for our desired quantile levels $\alpha(i,n)$.

An approximate solution is provided by Abramovitz and Stegun in Section 26.5.22 . It is easy to get an exact solution for the special case $a = n$. Then $p = .5^{(1/n)}$. Symmetry gives the case

$a = n$ as $1 - .5^{(1/n)}$. For $n = 9$ the exact answer is .0741... , but Abramovitz's approximation is .1034... . The Abramovitz approximation turns out to be much better in the middle of the interval than at the ends. Iterative improvement would seem to be worthwhile, particularly at the ends of the interval.

Newton iterative improvement worked very well with two iterations generally seeming to provide full word precision. The disadvantage of interative improvement is that one iteration to improve a single p_i requires i arithmetic steps. This might be annoying for large N as it would imply N^2 computations to get all the p_i for a given N . The best way to handle the problem seemed to be to fix a desired threshold, 'THRESH', in our case .5/2112 since our plotter has 2112 distinguishable positions across the width of the paper. Then begin on the ends of the interval using iterative improvement. As the middle of the interval is approached, corrections to the Abramovitz solution become less and less, so that when one iteration implies less change than the required threshold, iterative improvement is abandoned from then on to the middle of the interval. Symmetry provides the other half of the interval.

The iterative improvement begins with the power series definition of the incomplete beta function, equation (1). Each term is obtained by a modification of the previous term. Also, a new variable $q = 1 - p$ is defined (note $dq = - dp$) , and the partial derivative of each term with respect to p and with respect to q is determined analytically. The final correction dp is obtained by solving

$$\frac{1}{2} = I_p + \frac{\partial I}{\partial p}\, dp + \frac{\partial I}{\partial q}\, dq$$

$$= I_p + \left(\frac{\partial I}{\partial p} - \frac{\partial I}{\partial q} \right) dp$$

$$dp = \frac{.5 - I_p}{\frac{\partial I}{\partial p} - \frac{\partial I}{\partial q}} \qquad (2)$$

A subroutine to do the job is in Figure 1. A test program with some results is in Figure 2.

```
      SUBROUTINE IIB(N,P)
C     FIND INVERSE INCOMPLETE BETA FUNCTION OF .5, THAT IS,
C     SOLVE FOR  (P(IA),IA=1,N)
C
C        N    N  S      N-S
C       SUM ( )P  (1-P)      = IBETA (A,N-A+1) = IBETA (A,B) =  .5
C       S=A  S                      P                    P
C
C     METHOD:
C     START WITH ABRAMOVITZ APPROXIMATION 26.5.22
C     USES EXACT VALUE AT IA=N
C     ITERATIVE IMPROVEMENT UNTIL CHANGE LESS THAN 'THRESH' OR 4 ITERS
      DIMENSION P(N)
      THRESH=.5/2112
      P(N)=.5**(1./N)
      IF(N.EQ.1) RETURN
      NH=(N+1)/2
      DO 10 IB=2,NH
      IA=N-IB+1
      AINV=1./(2.*IA-1.)
      BINV=1./(2.*IB-1.)
      H=2./(AINV+BINV)
      W=(AINV-BINV)*(-.5+5./6.-2./(3.*H))
   10 P(IA)=IA/(IA+IB*EXP(2.*W))
      DO 40 IB=2,NH
      DO 30 ITER=1,4
      IA=N-IB+1
      Q=1.-P(IA)
      T=P(IA)**N
      BETA=T
      TP=N*T/P(IA)
      TQ=0.
      DO 20 IS=2,IB
      T=T*(N-IS+2)*Q/((IS-1)*P(IA))
      TP=TP+T*(N-IS+1)/P(IA)
      TQ=TQ+T*(IS-1)/Q
   20 BETA=BETA+T
      DP=(.5-BETA)/(TP-TQ)
      P(IA)=P(IA)+DP
      IF(ABS(DP).LT.THRESH.AND.ITER.EQ.1) GO TO 50
   30 IF(ABS(DP).LT.THRESH) GO TO 40
   40 CONTINUE
   50 DO 60 I=1,NH
   60 P(I)=1.-P(N-I+1)
      RETURN
      END
```

Figure 1. Fortran subroutine to find quantiles estimated by N ordered realizations of a random process.

```
C       TEST IBETA
        DIMENSION P(10)
        DO 10 N=1,9
        CALL IIB(N,P)
 10     PRINT 20,(P(I),I=1,N)
 20     FORMAT(10F8.5)
        STOP
        END
```

```
0.50000
0.29289 0.70711
0.20630 0.50000 0.79370
0.15910 0.38573 0.61427 0.84090
0.12945 0.31381 0.50000 0.68619 0.87055
0.10910 0.26445 0.42141 0.57859 0.73555 0.89090
0.09428 0.22849 0.36412 0.50000 0.63588 0.77151 0.90572
0.08300 0.20113 0.32052 0.44016 0.55984 0.67948 0.79887 0.91700
0.07413 0.17962 0.28624 0.39308 0.50000 0.60692 0.71376 0.82038 0.92587
```

Figure 2. A test program for the inverse incomplete beta function of one half subroutine IIB in Figure 1. Note that for $N = 9$ the quantiles are approximately $i/10$.