



## Algorithm AS R86: A Remark on Algorithm AS 152: Cumulative Hypergeometric Probabilities

Roger L. Berger

*Applied Statistics*, Vol. 40, No. 2 (1991), 374-375.

Stable URL:

<http://links.jstor.org/sici?sici=0035-9254%281991%2940%3A2%3C374%3AAARARO%3E2.0.CO%3B2-I>

*Applied Statistics* is currently published by Royal Statistical Society.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/rss.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

```

      SUM = SUM + V(J * (J - 1) / 2 + I) * G(J)
180    CONTINUE
      WORK(K + I) = SUM + V(I * (I + 1) / 2) * G(I)
190    CONTINUE
C
C      Update the ALPHAs
C
      NITER = ITN
      DO 200 J = 1, K
        ALPHA(J) = ALPHA(J) + WORK(K + J)
        IF (ALPHA(J) .LE. ZERO) THEN
          IFAULT = 6
          RETURN
        END IF
200    CONTINUE
C
C      Test for convergence
C
      S = ZERO
      DO 210 J = 1, K
        S = S + G(J) * WORK(K + J)
210    CONTINUE
C
      IF (S .LT. CHI2) GO TO 230
C
220 CONTINUE
C
      IFAULT = 7
C
C      Note that GAMMAD cannot fail
C
230 EPS = GAMMAD(S / TWO, RK / TWO, IF1)
      RLOGL = ZERO
      SUM = ZERO
      DO 240 J = 1, K
        SUM = SUM + ALPHA(J)
        RLOGL = RLOGL + (ALPHA(J) - ONE) * WORK(J) -
          * AN * ALOGAM(ALPHA(J), IF1)
240 CONTINUE
      RLOGL = RLOGL + AN * ALOGAM(SUM, IF1)
C
      END

```

### Algorithm AS R86

#### A Remark on Algorithm AS 152: Cumulative Hypergeometric Probabilities

By Roger L. Berger†

*North Carolina State University, Raleigh, USA*

[Received March 1990: Revised June 1990]

Shea (1989) has provided an improved version of the algorithm CHYPER of Lund (1980) for calculating point and cumulative probabilities for the hypergeometric distribution. The following two additions will improve the algorithm's accuracy when computing the cumulative probability for a large value of the parameter LL.

The first recommended change is to insert the statement

†Address for correspondence: Statistics Department, North Carolina State University, Box 8203, Raleigh, NC 27695-8203, USA.

IF (LL .GT. REAL(NN\*KK + NN + KK + 1)/REAL(MM + 2)) CHYPER = ONE

between the statements

IFault = 3  
RETURN

This RETURN is taken if underflow occurs when computing

$$p(LL|KK, MM, NN) = \binom{NN}{LL} \binom{MM - NN}{KK - LL} / \binom{MM}{KK}.$$

In Shea's version, CHYPER always returns a value of 0.0 for the cumulative probability if underflow occurs. But if LL is in the upper tail of the hypergeometric distribution, near the maximum possible value  $b = \min(KK, NN)$ , then the cumulative probability is near 1.0. This is because the value of  $p(i|KK, MM, NN)$  is near 0.0 for all  $i > LL$ , since underflow occurred when computing  $p(LL|KK, MM, NN)$  and  $p(i|KK, MM, NN) < p(LL|KK, MM, NN)$ . Recall that

$$p(i-1|k, m, n) = \frac{i(m-n-k+i)}{(n-i+1)(k-i+1)} p(i|k, m, n).$$

The coefficient on  $p(i|k, m, n)$  is greater than unity and  $p(i|k, m, n) < p(i-1|k, m, n)$  if and only if  $i > (nk + n + k + 1)/(m + 2)$ . The suggested addition checks to see whether LL satisfies this inequality to determine whether LL is in the upper tail of the distribution. The underflow flag, IFAULT = 3, is still returned but the value of CHYPER = 1.0 should be accurate. Shea (1989) states that if IFAULT = 3 the probability returned may be in error. But with this change, if IFAULT = 3, the probability returned should always be accurate, despite the underflow.

The second recommended change concerns the special value  $LL = b$ , the maximum possible value. CHYPER should return the value 1.0 if the cumulative probability of  $LL = b$  is requested. To accomplish this, after the statement

IF (K .EQ. 1 .OR. K .EQ. M .OR. N .EQ. 1 .OR. N .EQ. M) RETURN

insert the statement

IF (.NOT. POINT .AND. LL .EQ. MIN(KK, NN)) RETURN

This will avoid the following two possible situations. For large values of MM, CHYPER uses a normal approximation for the cumulative probability and can return a value less than 1.0 for  $LL = b$ . Also, underflow can occur when CHYPER computes  $p(LL|KK, MM, NN)$ . Using the first recommended change, CHYPER will return the correct value of 1.0, but with IFAULT = 3. Some may find it wasteful to include this second recommended change. Even without it, the value returned by CHYPER should be accurate to the three-place accuracy mentioned by Shea (1989). But this second change allows CHYPER to return the correct value of 1.0 quickly, with IFAULT = 0 indicating that this is the correct value.

### References

- Lund, R. E. (1980) Algorithm AS 152: Cumulative hypergeometric probabilities. *Appl. Statist.*, **29**, 221-223.  
Shea, B. L. (1989) Remark AS R77—A remark on Algorithm AS 152: Cumulative hypergeometric probabilities. *Appl. Statist.*, **38**, 199-204.