

ON LITTLEWOOD'S ESTIMATE FOR THE BINOMIAL DISTRIBUTION

BRENDAN D. MCKAY,* *The Australian National University*

Abstract

We correct a theorem of J. E. Littlewood which gives an approximation for the tail of the binomial distribution. We also present several new approximations which are less accurate but have wider scope. One of them gives an estimate with relative error uniformly $O(1/\sigma)$ over all values of all the parameters, where σ is the standard deviation.

For some types of probability calculations, the familiar DeMoivre–Laplace approximation to the binomial distribution (see Feller [2] for example), is insufficiently accurate. The most serious attack on this problem seems to have been that of Littlewood [3]. Unfortunately, Littlewood's Theorem 2 contains two typographic errors as well as an incorrect sign which can be traced to a clerical error in the proof. The main purpose of this note is to state Littlewood's theorem correctly. We also take the opportunity to give some other approximations which may be more convenient, though lesser in accuracy and scope in some cases.

For the normal distribution, define $\phi(x) = e^{-x^2/2}/\sqrt{2\pi}$, $Q(x) = \int_x^\infty \phi(u) du$, and $Y(x) = Q(x)/\phi(x)$. For the binomial distribution, define

$$b(k) = b(k; n; p) = \binom{n}{k} p^k q^{n-k}$$

and

$$B(k) = B(k; n, p) = \sum_{j=k}^n b(j; n, p),$$

where $q = 1 - p$. The mean of this distribution is $\mu = pn$ and the variance is $\sigma^2 = npq$.

We begin with Littlewood's Theorem 2. With the errors corrected, we can state that theorem as follows.

Theorem 1. Let p , $0 < p < 1$, be fixed. Let $t = t(n)$ be such that $pn + t$ is an integer and $0 \leq t \leq \frac{3}{4}qn$. Define $x = t/\sigma$ and $\rho = q - t/n$. Then

$$B(pn + t; n, p) = Q(x) \exp(A_1 + A_2/\sqrt{\rho(1-\rho)n} + O(n^{-1})),$$

where

$$A_1 = \frac{t^2}{2pqn} - (pn + t - \frac{1}{2}) \log\left(1 + \frac{t}{pn}\right) - (qn - t + \frac{1}{2}) \log\left(1 - \frac{t}{qn}\right)$$

and

$$A_2 = \frac{1}{6}(1 - 2\rho)((1 - x^2)/Y(x) + x^3) + \frac{1}{2}(1/Y(x) - x).$$

Proof. Littlewood's statement of this theorem has three errors:

- (i) The coefficient $\frac{1}{2}$ should be $\frac{1}{6}$ (it comes from c_3).
- (ii) The definition of ρ should be as in (his) Theorem 1.
- (iii) The sign of the $O(n^{-1/2})$ term is wrong.

Received 19 September 1988; revision received 13 December 1988.

* Postal address: Computer Science Department, Australian National University, GPO Box 4, ACT 2601, Australia.

Errors (i) and (ii) are merely typos. Error (iii) can be traced to an incorrect sign change in passing from (22.5) to (22.6). Apart from this, Littlewood's proof appears valid. To guard against other gross errors, we have successfully checked the theorem numerically for a wide range of values of the parameters.

Since $Y(x)$ has an asymptotic expansion of the form $1/x - 1/x^3 + O(1/x^5)$, the coefficient A_2 is uniformly bounded over the range of validity of Theorem 1. In fact, $0 < A_2 < 0.532$ and $0 < (1+x)A_2 < 1.084$ for $x \geq 0$ and $0 \leq \rho \leq 1$. Also, as Littlewood makes clear in his paper, the choice of $\frac{3}{4}qn$ as the upper limit for t is arbitrary; in fact the theorem and its proof are equally valid for $0 \leq t \leq \alpha qn$, where α is any constant with $\alpha < 1$. Subject to this bound on t , the error term in Theorem 1 is uniform over n and t . However, it is not uniform over p or α .

We note here that the slip leading to error (iii) also invalidates Littlewood's Theorem 1, but in that case we have not determined the correct form.

Littlewood's method (but not his theorem) would doubtless work if p was not constant but instead decreased as some function of n . We have not attempted to modify his proof in this way, but have instead taken several alternative approaches.

The first approach is suggested by Littlewood's theorem. The second and third terms of A_1 strongly suggest Stirling's approximation to $\log b(pn + t)$. Applying this approximation in reverse, and dropping A_2 , gives the estimate $B(pn + t) \approx \sigma b(pn + t)Y(x)(1 + t/(pn))$. Amazingly, the error in this approximation turns out to be uniformly $O(1/\sigma)$ for all n, p and $t \geq 0$.

Theorem 2. Let $0 < p < 1, n \geq 1$, and $pn \leq k \leq n$. Define $x = (k - pn)/\sigma$. Then

$$B(k; n, p) = \sigma b(k - 1; n - 1, p)Y(x) \exp(E(k; n, p)/\sigma),$$

where

$$0 \leq E(k; n, p) \leq \min \{ \sqrt{\pi/8}, 1/x \}.$$

Proof. Define $B^*(k) = \sigma b(k - 1; n - 1, p)Y(x)$ and $b^*(k) = B^*(k) - B^*(k + 1)$. We bound $E(k) = E(k; n, p)$ by comparing $B^*(n)$ with $B(n)$ and $b^*(k)$ with $b(k)$ for $k < n$.

We begin by recalling some standard properties of the function $Y(x)$. For any $x \geq 0$, the sequence $Y(x), Y'(x), Y''(x), \dots$, alternates in sign. With the help of the identity $Y'(x) = xY(x) - 1$, this fact yields many inequalities for $Y(x)$ and its derivatives. For example, the inequalities $Y' < 0$ and $Y'' > 0$ imply that $x/(1 + x^2) < Y(x) < 1/x$ for $x > 0$.

First consider the case $k = n$. From the definitions, we find that $B^*(n) = B(n)xY(x)$, where $x = \sqrt{qn/p}$. Thus, $E(n) = -x \ln(xY(x))$, from which standard methods yield the inequalities $0 \leq E(n) \leq \min \{ \frac{1}{2}, 1/x \}$.

Next, consider $0 \leq k \leq n - 1$. Define $e(k) = e(k; n, p)$ by $b(k) = b^*(k) \exp(e(k)/\sigma)$. After a little algebra, we find that $b^*(k) = b(k)f(x)$, where

$$f(x) = (\sigma + qx)Y(x) - (\sigma - px)Y(x + 1/\sigma).$$

Since $k < n$, we have $\sigma - px > 0$. Also, since $Y'' > 0$, $Y(x + 1/\sigma) > Y(x) + Y'(x)/\sigma$. This yields $f(x) < 1 + pxY'(x)/\sigma \leq 1$, which implies that $e(k) \geq 0$.

Similarly, for $x > 0$,

$$\begin{aligned} f(x) &= \sigma(Y(x) - Y(x + 1/\sigma)) + x(pY(x) + qY(x + 1/\sigma)) \\ &\geq -Y'(x + 1/\sigma) + x(pY(x) + qY(x + 1/\sigma)) \\ &= 1 - \frac{1}{\sigma} Y(x + 1/\sigma) + qx(Y(x) - Y(x + 1/\sigma)) \\ &\geq 1 - \frac{1}{\sigma} Y(x + 1/\sigma) \\ &\geq \frac{\alpha x}{\alpha x + 1}, \end{aligned}$$

from which it follows that $e(k) \leq 1/x$.

To obtain the uniform bound, define $u = x + 1/(2\sigma)$. Then

$$\begin{aligned} f(x) &= \sigma(Y(x) - Y(x + 1/\sigma)) + x(pY(x) + qY(x + 1/\sigma)) \\ &\geq -Y'(u) + x(pY(x) + qY(x + 1/\sigma)) \\ &\geq 1 - \frac{1}{2\sigma} Y(u) - x(Y(u) - Y(x + 1/\sigma)) \\ &= 1 - \frac{1}{2\sigma} Y(u) + \frac{x}{2\sigma} Y'(u). \end{aligned}$$

The derivative of the last expression with respect to x is $xY''(u)/(2\sigma) \geq 0$, so the smallest value is achieved at $x = 0$. Hence

$$f(x) \geq 1 - \frac{1}{2\sigma} Y\left(\frac{1}{2\sigma}\right),$$

from which the bound $e(k) \leq \sqrt{\pi/8}$ follows easily.

Finally, we note that our bounds on $e(k)$ are non-increasing for $x \geq 0$, implying that they hold also for $E(k)$.

An estimate similar to Theorem 2 was obtained by Bahadur [1], although the uniformity of the estimate was not established.

Theorem 2 could probably be sharpened to $O(1/\sigma^2)$ without a great deal of effort. For a smaller range of k values, we have found such an estimate of quite different form.

Theorem 3. Let $p = p(n)$ be such that $\sigma \rightarrow \infty$ as $n \rightarrow \infty$. Let $t = t(n)$ be such that $pn + t$ is an integer and $t = O(\sigma^3)$. Then $B(np + t; n, p) = Q(t/\sigma + \Delta + O(\sigma^{-2} + |t|^5/\sigma^9))$, where

$$\Delta = -\frac{1+p}{3\sigma} + \frac{(5-11p-p^2)t}{36\sigma^3} - \frac{(1-2p)t^2}{6\sigma^3} + \frac{(5-14pq)t^3}{72\sigma^5} - \frac{(1-2p)(83-146pq)t^4}{2160\sigma^7}.$$

Proof. Define $h(t) = h(t, n, p)$ by $B(np + t) = Q(t/\sigma + h(t))$. With a simple iteration, we can find a function $h_1(t)$ such that, for $t = O(\sigma^3)$,

$$(1) \quad Q(t/\sigma + h_1(t)) - Q((t+1)/\sigma + h_1(t+1)) = (1 + O(\sigma^{-7/4} + |t|^3/\sigma^6 + |t|^5/\sigma^9))b(np + t).$$

There is such a function $h_1(t)$ of the form

$$h_1(t) = \Delta + \frac{\beta_4(p)t^5}{\sigma^9} + \frac{\beta_5(p)t^6}{\sigma^{11}} + \frac{\beta_6(p)t^7}{\sigma^{13}},$$

where each $\beta_i(p)$ is a polynomial in p of degree i .

Thus, putting $m = \lceil \sigma^{7/4} \rceil$,

$$\begin{aligned} &Q(t/\sigma + h_1(t)) - Q((t+m)/\sigma + h_1(t+m)) \\ &= B(np + t) - B(np + t + m) + \sum_{r=0}^{m-1} b(pn + t + r)O(\sigma^{-2} + |t+r|^3/\sigma^6 + |t+r|^5/\sigma^9). \end{aligned}$$

Assume now that $t \geq 0$. For integer $k \geq 0$,

$$\sum_{r=0}^k (t+r)^k b(pn + t + r) = b(np + t)O(\sigma^2(\sigma + t)^{k-1}).$$

Thus,

$$(2) \quad Q(t/\sigma + h_1(t)) = B(pn + t) + b(pn + t)O(1/(\sigma + t) + (\sigma + t)^2/\sigma^4 + (\sigma + t)^4/\sigma^7),$$

the other terms being easily seen to be negligible in comparison.

Now define $\epsilon(t) = h(t) - h_1(t)$. By the mean-value theorem applied to (1), we have $ob(pn + t) = \phi(t/\sigma + h_1(t))(1 + \sigma(1))$, and so Equation (2) gives

$$Q(t/\sigma + h_1(t)) - Q(t/\sigma + h_1(t) + \epsilon(t)) = \phi(t/\sigma + h_1(t))O(\sigma^{-2} + t^2/\sigma^5 + t^4/\sigma^8),$$

from which it easily follows that $\epsilon(t) = O(\sigma^{-2} + t^2/\sigma^5 + t^4/\sigma^8)$.

The theorem now follows for $t \geq 0$. For $t < 0$, we need only note that $B(np + t; n, p) = 1 - B(nq - t + 1; n, q)$.

The symbolic algebra system Maple [4] was used for these calculations. Both Theorems 2 and 3 could be extended to more terms or a wider range if required.

Acknowledgements

I wish to thank Nick Wormald for some useful ideas. Also, the referee deserves thanks for detecting an error in the first draft of this note, thereby (one hopes) obviating the need for a third iteration!

References

- [1] BAHADUR, R. R. (1960) Some approximations to the binomial distribution function. *Ann. Math. Statist.* **31**, 43–54.
- [2] FELLER, W. (1969) *An Introduction to Probability Theory and its Applications*, Vol. 1. Wiley, New York.
- [3] LITTLEWOOD, J. E. (1969) On the probability in the tail of a binomial distribution. *Adv. Appl. Prob.* **1**, 43–72.
- [4] MAPLE SYMBOLIC ALGEBRA SYSTEM, Version 4.1 (1987) Symbolic Computation Group, University of Waterloo.