TWO ESTIMATES OF THE BINOMIAL DISTRIBUTION

By J. C. Mott-Smith

Air Force Cambridge Research Laboratories

1. Introduction and summary. The tail of the binomial distribution is defined as

$$E(n, s, p) = \sum_{r=s}^{n} \binom{n}{r} p^r (1-p)^{n-r} \tag{1}$$

with the restriction that p be non-negative and less than (approximately) s/n. An estimate of $E(n, s, p)$ can be considered as the product of two separate estimates, an estimate of the size of the leading term in (1) and an estimate of the ratio of $E(n, s, p)$ to the leading term. This paper is concerned solely with the second estimate. That is, let

$$R = E(n, s, p) \Big/ \binom{n}{s} p^s (1-p)^{n-s}; \tag{2}$$

two estimates of R are presented.

In Section 2 an upper bound to R is derived from a geometric series approach. The resulting bound is an improvement over previous results in that it is useful even when p is near s/n, as opposed to simpler geometric bounds, such as that of Bahadur [1], which either blow up or become excessively large when p is near s/n. The bound is given in Theorem 1, Equation (9). Section 3 discusses the error in using this bound.

Section 4 gives a normal approximation to R, namely R^* of Equation (18). Theorem 3 shows that the relative error of this estimate goes to zero as s and $n - s$ go to infinity provided $0 \leqq p \leqq (s-1)/(n-1)$ and under a weak restriction on s/n in the limit. The uniformity of this result with p contrasts with the many normal approximations to $E(n, s, p)$, such as those of Bernstein, [2] Feller, [6], [7] Uspensky, [9] and Camp [3] in which the relative error becomes infinite as p approaches zero. Section 5 presents a brief discussion of the behavior of R^*.

2. A geometric bound on R. Define the following:

$$\alpha = (s+1)/(n+1) \tag{3}$$

$$z = p/(1-p) \tag{4}$$

$$x_1 = z(1-\alpha)/\alpha \tag{5}$$

$$k = [2(n+1)(1-\alpha)\alpha]^{\frac{1}{2}} \tag{6}$$

$$\gamma_r = (n-s+1-r)/(s+r). \tag{7}$$

Received 16 April 1963; revised 15 November 1963.

Note that $x_1 = z\gamma_1$. Let m be the integer defined by

$$k + 1 \geqq m > k. \tag{8}$$

The principal result of this section is

THEOREM 1. *Let* $0 \leqq s \leqq n - 1$, $n > 1$, *and* $0 < p \leqq \alpha$ *and let*

$$R_k = \left\{k + 2 - 2\alpha + \frac{2x_1}{1 - x_1}\left[1 - \frac{k+1-m}{k} x_1^m - \frac{x_1}{1-x_1} \cdot \frac{1 - x_1^{m-1}}{k}\right]\right\} \Big/ [2 + (1 - x_1)(k - 2\alpha)]. \tag{9}$$

Then $R < R_k$.

We remark on the excepted cases. Equality holds for $p = 0$, $0 \leqq s \leqq n - 1$, and for $n = 1$, $s = 0$ as can be easily verified. For $s = n$ equality may be considered to hold formally, for if (6) is substituted for k and m is set equal to 1, then we find that R_k approaches 1, which is the correct value.

PROOF OF THEOREM 1. Using the definitions of z and γ_r, we get for R

$$R = 1 + \sum_{r=1}^{n-s} z^r \prod_{i=1}^{r} \gamma_i . \tag{10}$$

Let $k > 1$ be an arbitrary positive number, ignoring expression (6) for the moment, but let m be given by (8) and γ_{k+1} by (7). Then (10) yields

$$\begin{aligned}(1 - z\gamma_{k+1})R &= 1 + (\gamma_1 - \gamma_{k+1})z + (\gamma_2 - \gamma_{k+1})\gamma_1 z^2 \\ &\quad + \cdots + (\gamma_m - \gamma_{k+1})\gamma_1 \cdots \gamma_{m-1} z^m \\ &\quad -\{(\gamma_{k+1} - \gamma_{m+1})\gamma_1 \cdots \gamma_m z^{m+1} + \cdots + (\gamma_{k+1} - \gamma_{n-s+1})\gamma_1 \cdots \gamma_{n-s} z^{n-s+1}\}.\end{aligned} \tag{11}$$

(Note that $\gamma_{n-s+1} = 0$ which insures that (11) is formally correct for all positive values of $k > 1$ no matter how large.)

The first series on the right side of (11) and the series within braces are positive (or zero) since

$$\gamma_r - \gamma_t = (n + 1)(t - r)/(s + t)(s + r) > 0 \qquad \text{for } t > r \geqq 0. \tag{12}$$

The bound is obtained from (11) by dropping the series in braces, bounding the remaining series above by a geometric series and then selecting an appropriate value for k. From (12), (11), and (5) it follows that

$$\begin{aligned}(1 - z\gamma_{k+1})R &\leqq 1 + \frac{n+1}{(n-s)(s+k+1)} \sum_{r=1}^{m} (k + 1 - r)x_1^r \\ &= 1 + \frac{n+1}{(n-s)(s+k+1)} \cdot \frac{x_1}{1-x_1} \cdot \left[k - (k + 1 - m)x_1^m - \frac{x_1}{1-x_1}(1 - x_1^{m-1})\right].\end{aligned} \tag{13}$$

Ideally k should be selected to minimize the bound for R for each value of x_1; however this can't be done analytically. Instead we treat the expected worst case and hope for the best: Put x_1 equal to 1 in (13) and further assume that k is an integer. Then in obvious notation

$$(14)\qquad R]_{x_1=1} \leqq [(n - s)/(n + 1)]\cdot[(s + k + 1)/k] + \tfrac{1}{2}(k + 1).$$

The right side of (14) is minimized when k is given by (6). Observe that for $0 \leqq s \leqq n - 1$ and $n > 1$ the condition $k > 1$ is satisfied; hence $m \geqq 2$ which is sufficient for strict inequality in (13). Combining these results gives (9).

3. Error bounds on R_k. A lengthy analysis is given in Reference [8] of the relative error $\mathcal{E}_k = (R_k - R)/R$, which is probably not of general interest. We confine ourselves here to a few remarks without proof.

REMARK 1. Although it has not been demonstrated, it appears that $\mathcal{E}_k$ decreases monotonically with p for fixed n and s. If this is true, the following theorem is useful for estimating the maximum error.

THEOREM 2. *If $p = \alpha$ and n and s tend to infinity so that α is bounded away from* 0 *and* 1, *then*

$$\lim_{n\to\infty} \mathcal{E}_k = 2/\pi^{\frac{1}{2}} - 1 \doteq 0.128.$$

Theorem 2 may be proved by applying Stirling's approximation to $\binom{n}{s} p^s(1 - p)^{n-s}$ and by observing that $E(n, s, p) \to \frac{1}{2}$ under the conditions of the theorem.

REMARK 2. There is no reason to presume from the proof of (9) that n need be large in order that $\mathcal{E}_k$ be small. Indeed, all the calculations which have been made exhibit the reverse behavior, so that asymptotic bounds which hold for large n and s are of interest. Such bounds and some calculations are presented in Mott-Smith [8].

A minor point is that (9) may be simplified in appearance somewhat by replacing m by $k + 1$ throughout. Although this decreases R_k slightly, the inequality $R < R_k$ will not be jeopardized unless k is very small, say smaller than 2 or 3.

Finally, the value of k was selected to minimize the bound for $p = \alpha$, which leaves open the question of how much has thus been lost when p is smaller than α. For comparison, consider the simple geometric bound of Bahadur [1] which takes the form: $R < 1/(1 - x_1)$ for $p > 0$, in the notation of this paper. Now, by setting the expression in brackets in the numerator of (9) equal to unity we have

$$(1 - x_1)R_k < [(k + 2 - 2\alpha)(1 - x_1) + 2x_1]/[2 + (1 - x_1)(k - 2\alpha)] = 1.$$

Thus, even with k given by (6), which is the optimum k only when $x_1 = 1$, R_k is better than Bahadur's bound for all x_1; the latter bound is known to be very close when x_1 is much smaller than unity.

4. A normal approximation to R. It is well known that $E(n, s, p)$ is proportional to a hypergeometric function whose integral representation leads to an expression for R [4]. With $z = p/(1-p)$, as before, we have

LEMMA 1.

$$(15) \qquad R = F(-n+s, 1; s+1; -z) = s\int_0^1 (1-t)^{s-1}(1+tz)^{n-s}\,dt.$$

The integrand of (15) is unity at $t = 0$, is zero at $t = 1$, and is monotonically decreasing in between if $0 \leqq p \leqq (s-1)/(n-1)$. This behavior suggests an approximation along the lines of the method of Laplace, by expanding the logarithm of the integrand in a Maclaurin series and dropping terms in t^3 and higher. The radius of convergence of this expansion is the smaller of 1 and $1/z$ which may be less than unity; however, even for large values of z the principal contribution to the integral arises within the radius of convergence. The following auxiliary definition will be used to prove the results:

$$(16a) \qquad a = 1 - b = (s-1)/(n-1),$$

$$(16b) \qquad h(t) = (1-t)^{s-1}(1+zt)^{n-s},$$

$$(16c) \qquad f(t) = \exp\left[-(n-1)\{(a-bz)t + (a+bz^2)t^2/2\}\right],$$

$$(16d) \qquad g(t) = h(t)/f(t).$$

Now, let

$$(17) \qquad R^* = s\int_0^1 f(t)\,dt.$$

Then

$$(18) \qquad R^* = \frac{s(2\pi)^{\frac{1}{2}}}{[s-1+(n-s)z^2]^{\frac{1}{2}}} \cdot \exp L^2/2 \cdot \int_L^U \exp(-x^2/2)(2\pi)^{-\frac{1}{2}}\,dx,$$

where

$L = [s-1-(n-s)z]/[s-1+(n-s)z^2]^{\frac{1}{2}}$ and

$$U = L + [s-1+(n-s)z^2]^{\frac{1}{2}}.$$

Let

$$(19) \qquad \mathcal{E}^* = (R^* - R)/R.$$

Our principal result is

THEOREM 3. *A sufficient condition that $\mathcal{E}^*$ go to zero as n goes to infinity is that s and $n - s$ go to infinity in such a way that* $(\ln s)/(n-s)^{\frac{1}{3}}$ *goes to zero, and* $0 \leqq p \leqq (s-1)/(n-1)$.

Some preliminaries to the proof are stated in the following lemmas.

LEMMA 2. *$f(t)$ is positive and monotonically decreasing if $p \leqq a$, that is, if $z \leqq a/b$.*

PROOF. By inspection $f(t)$ is positive and $\ln f(t)$ decreases with t for $z \leqq a/b$.

LEMMA 3. *Let* $b/a > t \geqq 0$ *and* $n - s > 0$. *Then for* $0 \leqq z \leqq a/b$

$$f(t) \leqq \exp(-(s-1)t^2/2b). \tag{20}$$

PROOF. Let n, s, t be fixed. The derivative of $\ln f(t)$ with respect to z is $(n - s)t(1 - zt)$ which is greater than zero so long as $t > 0$. The maximum therefore occurs at $z = a/b$ which gives the desired result upon substitution in (16c). If $t = 0, f(t) = 1$ and the assertion is trivially true.

LEMMA 4. *Let* $s > 2$ *and* $n - s > 1$. *Then* dh/dt *is negative unless either*: (i) $t = 1$ *or* (ii) $t = 0$ *and* $z = a/b$, *in which cases* $dh/dt = 0$.

PROOF. The derivative is $dh/dt = -(1 - t)^{s-2}(1 + zr)^{n-s-1}(n - 1)\cdot(a - bz + zt)$.

PROOF OF THEOREM 3. Let n, s, z be fixed for the moment and let $0 < t_n < 1$. Then from (16) and (19)

$$\mathcal{E}^* = (s/R)\int_0^1 (f - h)\,dt = (s/R)\int_0^{t_n} f(1 - g)\,dt + (s/R)\int_{t_n}^1 (f - h)\,dt.$$

There exists $0 \leqq c_n \leqq 1$ such that

$$|\mathcal{E}^*| < (s/R)|1 - g(c_nt_n)|\int_0^1 f\,dt + (s/R)[f(t_n) + h(t_n)](1 - t_n)$$
$$< |1 - g(c_nt_n)|(1 + |\mathcal{E}^*|) + sf(t_n)(1 + g(t_n)),$$

whence, for $g(c_nt_n) \neq 0$,

$$|\mathcal{E}^*| < [|1 - g(c_nt_n)| + sf(t_n)(1 + g(t_n))]/[1 - |1 - g(c_nt_n)|],$$

where we have used Lemmas 2 and 4 and the fact that $R \geqq 1$.

To prove that $|\mathcal{E}^*| \to 0$ it is sufficient to show that t_n can be chosen so that $g(t_n) \to 1$, $g(c_nt_n) \to 1$, and that $sf(t_n) \to 0$. We first show that for any choice of t_n, these three conditions will be satisfied if the four conditions

$$t_n/b < 1, \tag{21}$$

$$(s-1)t_n^3 \to 0, \tag{22}$$

$$(n-s)(at_n/b)^3 \to 0, \tag{23}$$

$$(s-1)t_n^2/2b - \ln s \to \infty \tag{24}$$

are satisfied.

Now $\ln g(t)$ may be bounded above by using the inequality

$$|\ln(1 + x)| < x^3/3(1 - x), \qquad |x| < 1.$$

We find

$$|\ln g(t)| < (s-1)t^3/3(1-t) + (n-s)z^3t^3/3(1-zt). \tag{25}$$

Inspection of (25) shows that $g(t_n) \to 1$ ($\ln g(t_n) \to 0$) provided (21), (22),

(23) all hold. Use is here made of the inequalities $z \leqq a/b < 1/b$. Now it is clear that if the sequence $\{t_n\}$ satisfies (21), (22), (23); so, too, does $\{t_n c_n\}$ and hence $g(t_n c_n) \to 1$. Finally, by application of Lemma 3, whose conditions are satisfied, we have that (24) implies that $sf(t_n) \to 0$.

To complete the proof of Theorem 3 we produce a sequence $\{t_n\}$ which satisfies (21) through (24). Choose

$$t_n = [4b \ln s/(s-1)]^{\frac{1}{2}}.$$

Upon substitution we obtain:

$$t_n/b = [4 \ln s/ab(n-1)^{\frac{1}{2}} \to 0,$$

$$(s-1)t_n^3 = [4b \ln s/(s-1)^{\frac{1}{3}}]^{\frac{3}{2}} \to 0,$$

$$(n-s)(at_n/b)^3 = [4a \ln s/(n-s)^{\frac{1}{3}}]^{\frac{3}{2}} \to 0, \tag{23$'$}$$

and

$$(s-1)t_n^2/2b - \ln s = \ln s \to \infty,$$

the limits all holding under the conditions of the theorem as $n \to \infty$. The most stringent condition and case is seen to be (23$'$) when $b \to 0$.

5. Discussion of R^*. The dependence of R^* on n, s, and p is not obvious from the form of (18). This dependence can be made more apparent, at least if p is not too close to s/n, by rewriting (18). Following standard notation, let $\varphi(x) = (2\pi)^{-\frac{1}{2}} \exp(-x^2/2)$. Then (18) may be written

$$R^* = s/[s - 1 - (n-s)z] \cdot \{L \textstyle\int_L^U \varphi(x)\, dx/\varphi(L)\}.$$

If L is large the expression in brackets approaches unity. We also remark that for all practical purposes U may as well be taken as infinity.

The proof of Theorem 3 does not illuminate the dependence of $\mathcal{E}^*$ on n. The upper bound on $|\mathcal{E}^*|$ is, however, uniform in p so that the following theorems are at least illustrative:

THEOREM 4. *If $p = 0$ then $\mathcal{E}^* \sim (2s+1)/(s-1)^3$ as $s \to \infty$.*

PROOF. If p $= 0$ then $R = 1$, $L = (s-1)^{\frac{1}{2}}$, $U = 2(s-1)^{\frac{1}{2}}$, whence, from (18) and (19)

$$\mathcal{E}^* = \frac{s}{(s-1)^{\frac{1}{2}}} \exp(-(s-1)/2) \int_{(s-1)^{\frac{1}{2}}}^{2(s-1)^{\frac{1}{2}}} \exp(-x^2/2)\, dx - 1. \tag{26}$$

The asymptotic expansion

$$\int_x^\infty \exp(-x^2/2)\, dx \sim \frac{\exp(-x^2/2)}{x}(1 - 1/x^2 + 1\cdot 3/x^4 - \cdots), \qquad x \to \infty,$$

applied to (26) yields the desired result.

THEOREM 5. *If n is odd, $s = (n+1)/2$, and $p = \frac{1}{2}$ then $\mathcal{E}^* \sim 3/4n$ as $n \to \infty$.*

PROOF. $\mathcal{E}^*$ may be written as

$$\mathcal{E}^* = \binom{n}{s} p^s(1-p)^{n-s} R^*/E(n, s, p) - 1. \tag{27}$$

Under the conditions of the theorem $E(n, s, p) = \frac{1}{2}$, $a = p = \frac{1}{2}$ whence, from (18)

$$R^* \sim \tfrac{1}{2}(\pi/2)^{\frac{1}{2}}[(n+1)/(n-1)^{\frac{1}{2}}][1 + O(\exp(-(n-1)/2/(n-1)^{\frac{1}{2}}))].$$

$\binom{n}{s}$ is obtained from Sterling's series (cf. Erdelyi [5] p. 47) which gives for $n!$

$$n! = (n/e)^n(2\pi n)^{\frac{1}{2}}(1 + 1/12n + O(1/n^2))$$

and for $\binom{n}{s}$, with $s = n(n+1)/2$

$$\binom{n}{s} = \left[\frac{2}{\pi n}\right]^{\frac{1}{2}}(1 - 3/4n + O(1/n^2)).$$

Combining these results with (27) proves the theorem.

As our last remark we sketch how a comparison between R^* and the various normal approximations to $E(n, s, p)$ can be made. It will be our object to show that under certain conditions $R \sim R^*$ but that the relative error of the normal approximation blows up. In order to avoid the trivial case that $R = 1$ if $p = 0$ we also require that $R \to \infty$. We therefore call R^* "nontrivial" if

$$R \sim R^* \to \infty, \qquad \text{as } n \to \infty. \tag{28}$$

A sufficient condition that (28) hold is that the conditions of Theorem 3 hold, that s/n be bounded away from 0 and 1, and that $s/n \sim p$. On the other hand, most normal approximations require not only that $s/n \sim p$ but also that $n^\epsilon(s/n - p) \to 0$ for some particular $1 > \epsilon > 0$. Although this is usually stated as part of a sufficiency condition, it, or one like it, is necessary as well. Thus it is possible to satisfy (28) while violating a necessary condition of the normal approximation by controlling how fast $s/n - p$ goes to zero. The proceeding argument can be applied, for instance, to the approximation of Feller [7] p. 178. We observe that $E(n, s, p) \to 0$ in the limit, so that R^* may be a computational improvement only for very small values of $E(n, s, p)$. The requirement, in the usual normal approximation, that $s/n - p$ not go to zero too slowly can be traced back to the estimate of $\binom{n}{s}p^s(1-p)^{n-s}$, which suggests that the success of our approach derives from dealing with R instead of with $E(n, s, p)$.

REFERENCES

[1] BAHADUR, R. R. (1960). Some approximations to the binomial distribution function. *Ann. Math. Statist.* **31** 43–54.

[2] BERNSTEIN, S. (1943). Retour au problème de l'èvaluation de l'approximation de la formule limite de Laplace. (In Russian). *Izvestia Akademic Nauk SSR. Serija matematičeskaja.* **7** 3–16. (From Feller (6)).

[3] CAMP, B. H. (1951). An approximation to the point binomial. *Ann. Math. Statist.* **22** 130–131.

[4] ERDELYI, A., et al. (1953). *Higher Transcendental Functions* **1** McGraw-Hill, New York.

[5] ERDELYI, A. (1956). *Asymptotic Expansions.* Dover.

[6] FELLER, WILLIAM (1945). On the normal approximation to the binomial distribution. *Ann. Math. Statist.* **16** 319–329.

[7] FELLER, WILLIAM (1950). *An Introduction to Probability Theory and Its Applications,* **1** Ch. VII. Wiley, New York.

[8] MOTT-SMITH, J. (1960). Estimation of the binomial distribution. Air Force Cambridge Laboratories Research Report, AFCRL-TR-60-367.

[9] USPENSKY, J. V. (1937). *Introduction to Mathematical Probability.* McGraw-Hill, New York.