[2] A. L. O'Toole, "On the system of curves for which the method of moments is the best method of fitting," *Annals of Math. Stat.*, Vol. 4 (1933), pp. 1–29.
[3] A. L. O'Toole, "A method of determining the constants in the bimodal fourth degree exponential function," *Annals of Math. Stat.*, Vol. 4 (1933), pp. 79–93.
[4] J. Ernest Wilkins, Jr., "A note on skewness and kurtosis," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 333–335.
[5] C. V. L. Charlier, "A new form of the frequency function," *Lund universitet, Acta*, N. F. Bd 24 (1928), Avd. 2, Art. no. 8, pp. 1–26.
[6] M. G. Kendall, *The Advanced Theory of Statistics*, Griffin & Co., London, Vol. II, p. 43.

AN APPROXIMATION TO THE BINOMIAL SUMMATION

By G. F. Cramer

Washington, D. C.

We consider the binomial expansion $(q + p)^n$, where $q = 1 - p$ and n is a positive integer. For given values of n, p, r, and s, where $np < r < s \leq n$, we are often interested in the probability $P(r \leq x \leq s)$ that the number of successes x will satisfy $r \leq x \leq s$.

When n does not exceed 50, we can use tables of the Incomplete Beta Function, or other convenient and accurate tables. For "large" values of n, we can use normal tables. When p is "small", we can use Poisson tables. However, it is often true that p is fairly small, and yet not small enough to give really accurate results when Poisson tables are employed in the usual way, while n is too large for use of the tables of the Incomplete Beta Function and yet too small for accurate use of normal tables.

It frequently happens that an upper bound for $P(r \leq x \leq s)$ would serve our purpose. We propose to show how to find this from Poisson tables with greater accuracy than could be obtained by using these tables in the ordinary way.

We shall denote the general term of the binomial expansion by $B_i = \binom{n}{i}p^i q^{n-i}$ and the general term of the corresponding Poisson distribution with the same value of p by $P_i = (pn)^i e^{-pn}/i!$. We shall also consider a second Poisson distribution whose general term is given by $P'_i = (p'n)^i e^{-p'n}/i!$, where $p' \neq p$ will be determined later.

We shall use the following notations:

$$U_i = B_{i+1}/B_i = (n - i)(p)/(i + 1)(1 - p); \tag{1}$$

$$V_i = P_{i+1}/P_i = pn/(i + 1); \tag{2}$$

$$V'_i = P'_{i+1}/P'_i = p'n/(i + 1); \tag{3}$$

$$U_i - V_i = p(np - i)/(i + 1)(1 - p). \tag{4}$$

From (4) we obtain at once the following:

Lemma I. *$U_i > V_i$ or $U_i < V_i$ according as $i < np$ or $i > np$.*

Thus, the size of the general term of the binomial expansion falls off more steeply to the right of $i = np$ than does that of the general Poisson term.

We can use lemma I to obtain an upper bound to $P(r \leq x \leq s)$ for any $r > np$. In fact,

$$\begin{aligned} B_r &= B_r P_r / P_r \,; \\ B_{r+1} &< B_r P_{r+1} / P_r \,; \\ B_{r+2} &< B_{r+1} P_{r+2} / P_{r+1} < B_r P_{r+2} / P_r \,; \\ &\cdot \\ &\cdot \\ &\cdot \\ B_s &< B_r P_s / P_r \,. \end{aligned}$$

Adding these, we obtain

$$P(r \leq x \leq s) = \sum_{i=r}^{s} B_i < (B_r/P_r) \sum_{i=r}^{s} P_i = (B_r/P_r)\left(\sum_{i=r}^{\infty} P_i - \sum_{i=s}^{\infty} P_i\right). \tag{5}$$

The quantity in parentheses in (5) can be found by use of the cumulative Poisson table provided, of course, it is within the range of that table, while the B_r/P_r can be computed directly.

In the work we have done so far, we have used a Poisson distribution which is less steep than the corresponding binomial distribution throughout the whole interval $np < r \leq x \leq n$. It seems reasonable to investigate the possibility of improving upon (5) by using a Poisson distribution having a different value p' in place of p, where p' is chosen so that the new Poisson distribution is of the same steepness at $x = r$ as is the binomial distribution. We wish to have $U_r = V'_r$ and $U_i \leq V'_i$ for all $r \leq i \leq n$. The first of these conditions requires that $(n - r)(p)/(r + 1)(1 - p) = p'n/(r + 1)$. Solving for p' we obtain

$$p' = (n - r)(p)/(n)(1 - p). \tag{6}$$

We are now ready to prove the following:

LEMMA II. *If p' is defined by* (6) *and if U_i, V_i, and V'_i are defined by* (1), (2), *and* (3) *respectively, then $U_i \leq V'_i < V_i$, provided $r > np$ and $i \geq r$.*

It is easy to see that $U_i/V'_i = (n - i)(p)(1 + i)/(1 + i)(1 - p)(np')$, and this can be reduced to $(n - i)/(n - r)$ by replacing p' by its value from (6). Then $U_i/V'_i \leq 1$ since $i \geq r$. Moreover, we have $V'_i/V_i = (p'n)(i + 1)/(i + 1)(pn) = p'/p = (n - r)/(n - np)$. But $r > np$ and hence $V'_i < V_i$. This completes the proof of Lemma II.

We are now in a position to obtain an inequality somewhat better than (5). The derivation of the new upper bound for $P(r \leq x \leq s)$ goes just as before except that each P_i is replaced by P'_i. We obtain the new inequality

$$P(r \leq x \leq s) < K' B_r / P'_r, \tag{7}$$

where $K' = \sum_{i=r}^{\infty} P'_i - \sum_{i=s}^{\infty} P'_i$.

We can get a lower bound as well as a somewhat improved upper bound for

$P(r \leq x \leq s)$ by calculating B_r and B_{r+1} directly and then applying (5) or (7) to find an upper bound M of $P(r + 1 \leq x \leq s)$. This gives the inequality

$$B_r + B_{r+1} < P(r \leq x \leq s) < B_r + M. \tag{8}$$

This could, of course, be still further improved by calculating directly still more of the B_i's and using a similar procedure, but one would not care to carry this very far.

To illustrate the various approximations, we have worked out a numerical example the results of which appear below. For convenience in checking, we have used a value of n which is within the range of the tables of the Incomplete Beta Function, even though we would ordinarily use our method only for larger values of n.

EXAMPLE. $s = n = 40$; $r = 10$; $p = 1/10$; $p' = 1/12$. The tables of the Incomplete Beta Function give $P(10 \leq x \leq 40) = .0050631$. Using Poisson tables in the usual way, we get $P(10, 4) - P(40, 4) = .008132$, which is not particularly good. Using inequality (5) we obtain: $B_{10}/P_{10} = .6790$ and $P(10 \leq x \leq 40) < .6790(.008132) = .005522$. Using (8) and calculating both B_{10} and B_{11}, we take $r = 11$ in the inequality (5) and obtain $B_{10} = .0035934$, $B_{11} = .0010889$, $P(11, 4) - P(40, 4) = .002840$, $B_{11}/P_{11} = .5657$, and hence $.004682 < P(10 \leq x \leq 40) < .003594 + .001607 = .00520$. Again using method (8), but calculating B_{12} also and using $r = 12$ in inequality (5), we get $.004974 < P(10 \leq x \leq 40) < .005099$, which is quite good. We can obtain a still better result by using inequality (7) instead of (5). Then $p' = 1/12$, $np' = 10/3$, $B_{10}/P'_{10} = 2.150 +$, $P(10, 10/3) - P(40, 10/3) = .002366$, and $P(10 \leq x \leq 40) < .005087$.