SOME INEQUALITIES AMONG BINOMIAL AND POISSON PROBABILITIES

T. W. ANDERSON[1]
COLUMBIA UNIVERSITY
and
S. M. SAMUELS[2]
PURDUE UNIVERSITY

1. Introduction

The binomial probability function

$$
\begin{aligned}
(1.1)\qquad b(k; n, p) &= \binom{n}{k} p^k(1-p)^{n-k}, & k &= 0, 1, \cdots, n,\\
&= 0, & k &= n+1, \cdots,
\end{aligned}
$$

can be approximated by the Poisson probability function

$$(1.2)\qquad p(k; \lambda) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, \cdots,$$

for $\lambda = np$ if n is sufficiently large relative to λ. Correspondingly, the binomial cumulative distribution function

$$(1.3)\qquad B(k; n, p) = \sum_{j=0}^{k} b(j; n, p), \qquad k = 0, 1, \cdots,$$

is approximated by the Poisson cumulative distribution function

$$(1.4)\qquad P(k; \lambda) = \sum_{j=0}^{k} p(j; \lambda), \qquad k = 0, 1, \cdots,$$

for $\lambda = np$. In this paper it is shown that the error of approximation of the binomial cumulative distribution function $P(k; np) - B(k; n, p)$ is positive if $k \leq np - np/(n+1)$ and is negative if $np \leq k$. In fact, $B(k; n, \lambda/n)$ is monotonically increasing for all n $(\geq \lambda)$ if $k \leq \lambda - 1$ and for all $n \geq k/(\lambda - k)$ if $\lambda - 1 < k < \lambda$, and is monotonically decreasing for all n $(\geq k)$ if $\lambda \leq k$. Thus

[1] Research supported by the Office of Naval Research under Contract Number Nonr-4259(08), Project Number NR 042-034. Reproduction in whole or in part permitted for any purpose of the United States Government.

[2] Research supported in part by National Science Foundation Grant NSF-GP-3694 at Columbia University, Department of Mathematical Statistics, and in part by Aerospace Research Laboratories Contract AF 33(657)11737 at Purdue University.

for most practical purposes the Poisson approximation overestimates tail probabilities, and the margin of overestimation decreases with n.

The probability function $b(k; n, \lambda/n)$ increases with n [to $p(k; \lambda)$] if $k \leq \lambda + \frac{1}{2} - (\lambda + \frac{1}{4})^{1/2}$ or if $\lambda + \frac{1}{2} + (\lambda + \frac{1}{4})^{1/2} \leq k$. These facts imply that for given n and λ $P(k; \lambda) - B(k; n, \lambda/n)$ increases with respect to k, to $\lambda + \frac{1}{2} - (\lambda + \frac{1}{4})^{1/2}$ and decreases with respect to k from $\lambda + \frac{1}{2} + (\lambda + \frac{1}{4})^{1/2}$.

When the Poisson distribution is used to approximate the binomial distribution for determining significance levels, in nearly all cases the actual significance level is less than the nominal significance level given by the Poisson, and the probability of Type I error is overstated. Similarly, the actual confidence level of confidence limits based on the Poisson approximation is greater than the nominal level. Section 4 gives precise statements of these properties.

In section 5 another approach is developed to the monotonicity of $B(k; n, \lambda/n)$ and $b(k; n, \lambda/n)$. Although the results are not as sharp as those in sections 2 and 3, the methods are interesting.

Throughout the paper we assume $\lambda > 0$ and $n \geq \lambda$ (or $1 \geq \lambda/(n-1)$ when used as a binomial parameter).

2. Inequalities among cumulative distribution functions

2.1. *A general inequality.* In this section we show that $B(k; n, \lambda/n)$ increases with n to $P(k; \lambda)$ if k is small relative to λ and decreases to $P(k; \lambda)$ if k is large.

For this purpose the following theorem of Hoeffding will be needed.

THEOREM (Hoeffding [1]). *Let $F(k)$ be the probability of not more than k successes in n independent trials where the i-th trial has probability p_i of success. Let $\lambda = p_1 + p_2 + \cdots + p_n$. Then*

$$\begin{aligned} &B(k; n, \lambda/n) \geq F(k), && k \leq \lambda - 1, \\ &B(k; n, \lambda/n) \leq F(k), && k \geq \lambda. \end{aligned} \tag{2.1}$$

Equality holds only if $p_1 = \cdots = p_n = \lambda/n$.

It is possible to obtain from this the following result.

THEOREM 2.1.

$$\begin{aligned} &B(k; n, \lambda/n) > B(k; n-1, \lambda/(n-1)), && k \leq \lambda - 1, \\ &B(k; n, \lambda/n) < B(k; n-1, \lambda/(n-1)), && k \geq \lambda. \end{aligned} \tag{2.2}$$

PROOF. If we choose $p_1 = 0$, $p_2 = \cdots = p_n = \lambda/(n-1)$, then $F(k) = B(k; n-1, \lambda/(n-1))$, and the result follows from Hoeffding's theorem.

COROLLARY 2.1.

$$\begin{aligned} &P(k; \lambda) > B(k; n, \lambda/n), && k \leq \lambda - 1, \\ &P(k; \lambda) < B(k; n, \lambda/n), && k \geq \lambda. \end{aligned} \tag{2.3}$$

Alternatively, we may take $p_1 = 1$, $p_2 = \cdots = p_n = (\lambda - 1)/(n-1)$, in which case $F(k) = B(k-1; n-1, (\lambda-1)/(n-1))$. Hence we have the next theorem.

THEOREM 2.2.

$$
\begin{aligned}
&B(k; n, \lambda/n) > B(k-1; n-1, (\lambda-1)/(n-1)), \qquad k \le \lambda - 1,\\
&B(k; n, \lambda/n) < B(k-1; n-1, (\lambda-1)/(n-1)), \qquad k \ge \lambda.
\end{aligned} \tag{2.4}
$$

The limits of (2.4) as $n \to \infty$ give the following corollary.

COROLLARY 2.2.

$$
\begin{aligned}
&P(k; \lambda) \ge P(k-1; \lambda-1), \qquad k \le \lambda - 1,\\
&P(k; \lambda) \le P(k-1; \lambda-1), \qquad k \ge \lambda.
\end{aligned} \tag{2.5}
$$

This corollary will be used in section 4. If λ increases by steps of 1, corollary 2.2 indicates that the probabilities in the tails $[0, \lambda - a]$ and $[\lambda + c, \infty]$ are increasing with λ if $a \ge 1$ and $c \ge 1$.

2.2. *Monotonicity near the mean.* If λ is an integer, theorem 2.1 includes all values of k. If λ is not an integer, what happens at the value of k which is between $\lambda - 1$ and λ? We shall show that $B(k; n, \lambda/n)$ increases with n if n is sufficiently large.

THEOREM 2.3.

$$
\begin{aligned}
&B(k; n, \lambda/n) > B(k; n-1, \lambda/(n-1)), \quad k < \lambda, n \ge \lambda/(\lambda - k),\\
&B(k; n, \lambda/n) < B(k; n-1, \lambda/(n-1)), \quad k \ge \lambda.
\end{aligned} \tag{2.6}
$$

PROOF. The second inequality is part of theorem 2.1. The first inequality is stated for $nk/(n-1) \le \lambda$. It will follow from that inequality for $nk/(n-1) \le \lambda \le (n-1)k/(n-2)$ [lemma 2.2] and the following lemma.

LEMMA 2.1. *The function* $D(k; n, \lambda) = B(k; n, \lambda/n) - B(k; n-1, \lambda/(n-1))$ *has at most one sign change from negative to positive in the interval* $0 < \lambda \le n - 1$.

PROOF. The case $k = n - 1$ is true since $B(n-1; n-1, \lambda/(n-1)) \equiv 1$. Hence we consider only $k < n - 1$. Since $D(k; n, \lambda)$ is 0 for $\lambda = 0$ and positive for $\lambda = n - 1$, it suffices to show that the derivative has at most two sign changes, from negative to positive to negative. Now

$$
\frac{d}{d\lambda} D(k; n, \lambda) = b(k; n-2, \lambda/(n-1)) - b(k; n-1, \lambda/n), \tag{2.7}
$$

$$
\frac{b(k; n-2, \lambda/(n-1))}{b(k; n-1, \lambda/n)} = \frac{(n-1-k)(1 + 1/(n-1))^{n-1}}{(n-1-\lambda)(1 + 1/(n-1-\lambda))^{n-1-k}}. \tag{2.8}
$$

Then (2.8) is increasing for $\lambda < k + 1$ and decreasing for $\lambda > k + 1$. Hence (2.7) has the desired sign change property and the lemma is proved.

LEMMA 2.2. *If* $nk/(n-1) \le \lambda \le (n-1)k/(n-2)$, *then*

$$
B(k; n, \lambda/n) > B(k; n-1, \lambda/(n-1)). \tag{2.9}
$$

PROOF. In n independent trials with $n - 1$ of the probabilities of success equal to λ/n and the remaining probability equal to p, the probability of at most k successes is

$$
pB(k-1; n-1, \lambda/n) + (1-p)B(k; n-1, \lambda/n). \tag{2.10}
$$

This is a decreasing function of p and is equal to $B(k; n, \lambda/n)$ when $p = \lambda/n$. The value of p which makes (2.10) equal to $B(k; n-1, \lambda/(n-1))$ is

$$p^* = \frac{B(k, n-1, \lambda/n) - B(k, n-1, \lambda/(n-1))}{b(k, n-1, \lambda/n)} \tag{2.11}$$

$$= \frac{n-1-k}{(\lambda/n)^k(1-\lambda/n)^{n-1-k}} \int_{\lambda/n}^{\lambda/(n-1)} u^k(1-u)^{n-2-k}\, du.$$

It suffices, then, to show that under the hypothesis $p^* > \lambda/n$. If

$$\lambda \leq (n-1)k/(n-2), \tag{2.12}$$

then the integrand in (2.11) is increasing in u over the range of integration. Hence for $\lambda \leq (n-1)k/(n-2)$

$$p^* > \frac{(n-1-k)(\lambda/(n-1) - \lambda/n)(\lambda/n)^k(1-\lambda/n)^{n-2-k}}{(\lambda/n)^k(1-\lambda/n)^{n-1-k}} \tag{2.13}$$

$$= \frac{(n-1-k)\lambda}{(n-1)(n-\lambda)},$$

which is at least λ/n if

$$\lambda \geq nk/(n-1), \tag{2.14}$$

which proves the lemma.

Theorem 2.2 follows from lemma 2.2, because lemma 2.1 indicates that if (2.9) holds for a given value of λ, it holds for all larger values.

COROLLARY 2.3.

$$\begin{aligned} P(k; \lambda) &> B(k; n, \lambda/n), \qquad k < \lambda, n \geq k/(\lambda - k), \\ P(k; \lambda) &< B(k; n, \lambda/n), \qquad k \geq \lambda. \end{aligned} \tag{2.15}$$

2.3. *Special cases.* Lemma 2.1 implies that for each n there is a number λ_n such that $B(k; n+1, \lambda/(n+1)) - B(k; n, \lambda/n)$ is negative if $\lambda < \lambda_n$ and is positive if $\lambda > \lambda_n$, and theorem 2.3 indicates that $k < \lambda_n < k + k/n$. There is no simple expression for λ_n. Is it true that $\lambda_n - k \sim k/n$? The answer is no, as can be seen in the case $k = 1$. The first derivative of $B(1; n, \lambda/n)$ with respect to n approaches 0 as $n \to \infty$, and the second derivative is negative for all $n > \lambda$ if $\lambda \geq (\frac{4}{3})^{1/2}$, while if $\lambda < (\frac{4}{3})^{1/2}$ it is negative if

$$n \geq [\lambda^2 + \lambda(4 - 3\lambda^2)^{1/2}/[2(\lambda^2 - 1)]. \tag{2.16}$$

Equivalently, the second derivative is negative if $\lambda > n/(n^2 - n + 1)^{1/2} \sim 1 + 1/(2n)$. Thus for $k = 1$, $\lambda_n < n/(n^2 - n + 1)^{1/2}$.

For the case $k = n - 1$, we can explicitly evaluate λ_n. We have $\lambda_n = n + 2 + 1/n - (1 + 1/n)^{n+1}$ so that $\lambda_n - k$ approaches $3 - e \sim .282$ as $n \to \infty$. Thus λ_n is smaller than $n - 1 + (n-1)/n$, as given by our general result. In particular, if $k = 1$ and $n = 2$, $\lambda_n = \frac{9}{8}$, which is smaller than $2/\sqrt{3}$ as given by the preceding paragraph.

3. Inequalities among probability functions

Let

$$r(k; n) = b(k; n, \lambda/n)/b(k; n-1, \lambda/(n-1)), \tag{3.1}$$

$$d(k; n) = b(k; n, \lambda/n) - b(k; n-1, \lambda/(n-1)). \tag{3.2}$$

Then

$$r(k; n)/r(k-1; n) = (n+1-k)(n-1-\lambda)/[(n-k)(n-\lambda)], \tag{3.3}$$

which is less or greater than one according to whether k is less or greater than $\lambda + 1$. Since $d(0; n) > 0$ (by theorem 2.3), $d(n; n) > 0$, and $\sum_{k=0}^{n} d(k; n) = 0$, we have the following proposition.

PROPOSITION 3.1. *For suitable nonnegative integers a_n and c_n, depending on λ,*

$$\begin{aligned} d(k; n) &> 0, \qquad 0 \le k \le a_n \quad \text{or} \quad c_n \le k \le n, \\ d(k; n) &\le 0, \qquad a_n < k < c_n, \end{aligned} \tag{3.4}$$

and

$$\begin{aligned} &\max_{0 \le k \le n} [B(k; n, \lambda/n) - B(k; n-1, \lambda/(n-1))], \\ &\min_{0 \le k \le n} [B(k; n, \lambda/n) - B(k; n-1, \lambda/(n-1))] \end{aligned} \tag{3.5}$$

occur at $k = a_n$ and c_n, respectively.

The following theorem gives a lower bound for a_n and an upper bound for c_n.

THEOREM 3.1. *If $0 \le k \le \lambda + \frac{1}{2} - (\lambda + \frac{1}{4})^{1/2}$ or if $\lambda + \frac{1}{2} + (\lambda + \frac{1}{4})^{1/2} \le k \le n$, then*

$$b(k; n, \lambda/n) > b(k; n-1, \lambda/(n-1)); \tag{3.6a}$$

if $\lambda + \frac{1}{2} - (\lambda + \frac{1}{4})^{1/2} < k < \lambda + \frac{1}{2} + (\lambda + \frac{1}{4})^{1/2}$ and n is sufficiently large,

$$b(k; n, \lambda/n) < b(k; n-1, \lambda/(n-1)). \tag{3.6b}$$

PROOF. The conditions on k in the first part of the theorem are equivalent to $\lambda \le k - k^{1/2}$ and $\lambda \ge k + k^{1/2}$. Since

$$r(k; n) = b(k; n, \lambda/n)/b(k; n-1, \lambda/(n-1)) \tag{3.7}$$

decreases for $\lambda < k$ and increases for $\lambda > k$, it suffices to consider $b(k; n, \lambda/n)$ at only the values $\lambda = k \pm k^{1/2}$. We prove (3.6a) by showing that $\log b(k; n, \lambda/n)$ is an increasing function of n at these two values of λ. Now

$$\begin{aligned} \log b(k; n, \lambda/n) - \log \lambda^k + \log k! &= \sum_{j=1}^{k-1} \log(1 - j/n) + (n-k)\log(1 - \lambda/n) \\ &= -\sum_{j=1}^{k-1} \sum_{r=1}^{\infty} (j/n)^r/r - (n-k) \sum_{r=1}^{\infty} (\lambda/n)^r/r \\ &= -\lambda - \sum_{r=1}^{\infty} (1/n)^r \left[\sum_{j=1}^{k-1} j^r/r + \lambda^{r+1}/(r+1) - k\lambda^r/r\right]. \end{aligned} \tag{3.8}$$

Hence the first part of the theorem is proved if for $\lambda = k \pm k^{1/2}$, each bracketed expression is nonnegative. (Note that, if $k = 0$ or 1, the sums on j should be taken to be zero, and we see immediately that in these cases the theorem is true.)

For $r = 1$ the bracketed expression vanishes. In the general term, we replace the sum by a more manageable expression as follows: by convexity of z^r,

$$\int_x^{x+1} z^r \, dz \leq \tfrac{1}{2}[x^r + (x+1)^r]. \tag{3.9}$$

Hence,

$$\begin{aligned} \frac{(k-1)^{r+1}}{r+1} + \frac{(k-1)^r}{2} &= \int_0^{k-1} z^r \, dz + \tfrac{1}{2}(k-1)^r \\ &= \sum_{x=0}^{k-2} \int_x^{x+1} z^r \, dz + \tfrac{1}{2}(k-1)^r \\ &\leq \tfrac{1}{2} \sum_{x=0}^{k-2} [x^r + (x+1)^r] + \tfrac{1}{2}(k-1)^r \\ &= \sum_{j=0}^{k-1} j^r. \end{aligned} \tag{3.10}$$

Thus, each bracketed expression in (3.8) is at least

$$[(k-1)^{r+1} + \tfrac{1}{2}(r+1)(k-1)^r + r\lambda^{r+1} - (r+1)k\lambda^r]/[r(r+1)]. \tag{3.11}$$

Setting $\lambda = k + k^{1/2}$ and letting $u = k^{1/2}$, $v = k^{-1/2} \leq 1$, we can write (3.11) as

$$\begin{aligned} &[(u+1)^{r+1}(u-1)^{r+1} + \tfrac{1}{2}(r+1)(u+1)^r(u-1)^r + ru^{r+1}(u+1)^{r+1} \\ &\qquad - (r+1)u^{r+2}(u+1)^r]/[r(r+1)] \\ &= (u+1)^r\{(u-1)^r[u^2 + \tfrac{1}{2}(r-1)] - u^r(u^2 - ru)\}/[r(r+1)] \\ &= (u+1)^r u^{r+2}\{(1-v)^r[1 + \tfrac{1}{2}(r-1)v^2] - (1 - rv)\}/[r(r+1)] \\ &\geq 0, \end{aligned} \tag{3.12}$$

since $(1-v)^r \geq 1 - rv$ for $v \leq 1$, $r \geq 1$. A similar argument establishes the result at $\lambda = k - k^{1/2}$, which completes the proof of the first part of the theorem. The second part of the theorem follows since the coefficient of $1/n$ in (3.8) is positive for $k - k^{1/2} < \lambda < k + k^{1/2}$.

COROLLARY 3.1. *If $m > n$ and if $0 \leq k \leq \lambda + \frac{1}{2} - (\lambda + \frac{1}{4})^{1/2}$ or $\lambda + \frac{1}{2} + (\lambda + \frac{1}{4})^{1/2} \leq k \leq m$, then*

$$p(k; \lambda) > b(k; m, \lambda/m) > b(k; n, \lambda/n). \tag{3.13}$$

Hence,

$$\max_{0 \leq k \leq \lambda + \frac{1}{2} - (\lambda + \frac{1}{4})^{1/2}} [P(k; \lambda) - B(k; n, \lambda/n)] \tag{3.14}$$

occurs at the largest integer which is not greater than $\lambda + \frac{1}{2} - (\lambda + \frac{1}{4})^{1/2}$, and

$$\min_{\lambda + \frac{1}{2} + (\lambda + \frac{1}{4})^{1/2} \leq k \leq n} [P(k; \lambda) - B(k; n, \lambda/n)] \tag{3.15}$$

occurs at the smallest integer which is at least $\lambda + \frac{1}{2} + (\lambda + \frac{1}{4})^{1/2}$.

4. Applications to statistical inference

4.1. *Testing hypotheses.* Suppose n independent trials are made, each trial with probability p of success. Consider testing the null hypothesis $p = p_0$, where p_0 is specified. Against alternatives $p < p_0$ a (uniformly most powerful) nonrandomized test is a rule to reject the null hypothesis if the observed number of successes is less than or equal to an integer $\underline{k}$. The significance level of the test is $B(\underline{k}; n, p_0)$. It may be approximated by $P(\underline{k}; np_0)$. If $\underline{k} \leq np_0 - 1$, then by corollary 2.1, $B(\underline{k}; n, p_0) < P(\underline{k}; np_0)$. (Corollary 2.3 allows us to raise the bound on k to $np_0 - np_0/(n + 1)$, but we shall use here the simpler bound.) The procedure is conservative in the sense that the actual significance level (defined by the binomial distribution) is less than the nominal significance level (given by the Poisson distribution). The condition $\underline{k} \leq np_0 - 1$, which can be verified by the statistician in defining the test procedure, enables him to say that the probability of rejecting the null hypothesis when it is true is less than the approximating probability.

Against alternatives $p > p_0$ a (uniformly most powerful) nonrandomized test consists in rejecting the null hypothesis if the number of successes is greater than or equal to an integer $\bar{k}$. The significance level of the test is

$$1 - B(\bar{k} - 1; n, p_0), \tag{4.1}$$

which may be approximated by $1 - P(\bar{k} - 1; np_0)$. If $np_0 \leq \bar{k} - 1$, then by corollary 2.1, $1 - B(\bar{k} - 1; n, p_0) < 1 - P(\bar{k} - 1; np_0)$ and the procedure is conservative. A nonrandomized test against two-sided alternatives $p \neq p_0$ consists in rejecting the null hypothesis if the number of successes is less than or equal to $\underline{k}$ or greater than or equal to $\bar{k}$. The significance level $B(\underline{k}; n, p_0) + 1 - B(\bar{k} - 1; n, p_0)$ may be approximated by $P(\underline{k}; np_0) + 1 - P(\bar{k} - 1; np_0)$. If $\underline{k} + 1 \leq np_0 \leq \bar{k} - 1$, the procedure is conservative.

A (uniformly most powerful) randomized test of the null hypothesis $p = p_0$ against alternatives $p < p_0$ consists of a rule to reject the null hypothesis if the observed number of successes is less than $\underline{k}$ and to reject the null hypothesis with probability $\underline{\pi}$ if the number of successes is $\underline{k}$. The significance level is $\underline{\pi}B(\underline{k}; n, p_0) + (1 - \underline{\pi})B(\underline{k} - 1; n, p_0)$, which may be approximated by $\underline{\pi}P(\underline{k}; np_0) + (1 - \underline{\pi})P(\underline{k} - 1; np_0)$. The Poisson approximation overestimates the significance level if $\underline{k} \leq np_0 - 1$, since both $P(k; np_0)$ and $P(k - 1; np_0)$ overestimate the corresponding binomial probabilities. A (uniformly most powerful) randomized test against alternatives $p > p_0$ consists of a rule to reject the null hypothesis if the observed number of successes is greater than $\bar{k}$ and to reject the null hypothesis with probability $\bar{\pi}$ if the number of successes is $\bar{k}$. The significance level is $1 - \bar{\pi}B(\bar{k} - 1; n, p_0) - (1 - \bar{\pi})B(\bar{k}; n, p_0)$, which may be approximated by $1 - \bar{\pi}P(\bar{k} - 1; np_0) - (1 - \bar{\pi})P(\bar{k}; np_0)$. The approximation is an overestimate if $np_0 \leq \bar{k} - 1$. A two-sided randomized test consists in rejecting the null hypothesis if the number of successes is less than $\underline{k}$ or greater than $\bar{k}$, rejecting the null hypothesis with probability $\underline{\pi}$ if the number of successes is $\underline{k}$ and rejecting the null hypothesis with probability $\bar{\pi}$ if the number of successes is $\bar{k}$. The significance level, $\underline{\pi}B(\underline{k}; n, p_0) + (1 - \underline{\pi})B(\underline{k} - 1; n, p_0) + 1 - \bar{\pi}B(\bar{k} - 1; n, p_0) - (1 - \bar{\pi})B(\bar{k}; n, p_0)$ is overestimated by $\underline{\pi}P(k; np_0) + (1 - \underline{\pi})P(\underline{k} - 1; np_0) + 1 - \bar{\pi}P(\bar{k} - 1; np_0) - (1 - \bar{\pi})P(\bar{k}; np_0)$ if $\underline{k} + 1 \leq np_0 \leq \bar{k} - 1$.

Some criteria for satisfying conditions for overestimation by the approximation will be derived from the following theorems.

THEOREM 4.1. *If* $P(\ell; \lambda) \leq P(h; m)$, *then* $m + \ell - h \leq \lambda$, *for* $\ell = h, h + 1, \cdots, m = h + 1, h + 2, \cdots, h = 0, 1, \cdots$.

PROOF. $P(\ell; \lambda) \leq P(h; m)$ implies $P(\ell; \lambda) \leq P(\ell; \ell + m - h)$ because $P(h; m) \leq P(\ell; \ell + m - h)$ by corollary 2.2, and $P(\ell; \lambda) \leq P(\ell; \ell + m - h)$ implies $\ell + m - h \leq \lambda$ because $P(\ell; \nu)$ is a decreasing function of ν. [In fact $dP(\ell; \nu)/d\nu = -p(\ell; \nu)$.]

Theorem 4.1 with $\ell = k$, $h = 0$, and $m = 1$ states that if $P(k; \lambda) \leq P(0; 1) = e^{-1} \sim .3679$, then $k \leq \lambda - 1$, which implies that $B(k; n, \lambda/n)$ is increasing in n to $P(k; \lambda)$. The theorem with $\ell = k - 1$, $h = 0$, and $m = 2$ states that if $P(k - 1; \lambda) \leq P(0, 2) = e^{-2} \sim .1353$ then $k \leq \lambda - 1$, which implies that $B(k; n, \lambda/n)$ as well as $B(k - 1; n, \lambda/n)$ are increasing in n. Hence

$$\underline{\pi}P(\underline{k}; np_0) + (1 - \underline{\pi})P(\underline{k} - 1; np_0) \leq P(0; 2) \tag{4.2}$$

implies $P(\underline{k} - 1; np_0) \leq P(0; 2)$, which implies

$$\begin{aligned}\underline{\pi}B(\underline{k}; n, p_0) + (1 - \underline{\pi})B(\underline{k} - 1; n, p_0)& \\ < \underline{\pi}P(\underline{k}; np_0) + (1 - \underline{\pi})P(\underline{k}; np_0).&\end{aligned} \tag{4.3}$$

THEOREM 4.2. *If* $P(h; m) \leq P(\ell; \lambda)$, *then* $\lambda \leq m + \ell - h$, *for* $\ell = h, h + 1, \cdots, h = m, m + 1, \cdots, m = 0, 1, \cdots$.

PROOF. $P(h; m) \leq P(\ell; \lambda)$ implies $P(\ell; \ell + m - h) \leq P(\ell; \lambda)$ because $P(\ell; \ell + m - h) \leq P(h; m)$ by corollary 2.2, and $P(\ell; \ell + m - h) \leq P(\ell; \lambda)$ implies $\lambda \leq m + \ell - h$ because $P(\ell; \nu)$ is a decreasing function of ν.

Theorem 4.2 with $\ell = k - 1$ and $h = m = 1$ states that if $P(k - 1; \lambda) \geq P(1; 1) = 2e^{-1} \sim .7358$, then $\lambda \leq k - 1$; hence $1 - B(k - 1; n, \lambda/n)$ is increasing in n. In this case, $k = 2, 3, \cdots$. The test which rejects the null hypothesis on the basis of one or more successes ($\bar{k} = 1$) is not covered; in fact, $1 - B(0; n, \lambda/n)$ is decreasing in n. The theorem with $\ell = k$, $h = 2$, and $m = 1$ states that if $P(k; \lambda) \geq P(2; 1) = 5e^{-1}/2 \sim .9197$, then $\lambda \leq k - 1$, which implies that $B(k - 1; n, \lambda/n)$ as well as $B(k; n, \lambda/n)$ are decreasing in n. Hence

$$1 - \bar{\pi}P(\bar{k} - 1; np_0) - (1 - \bar{\pi})P(\bar{k}; np_0) \leq 1 - P(2; 1) \sim .0803 \tag{4.4}$$

implies $1 - P(\bar{k}; np_0) \leq 1 - P(1; 2)$, which implies

$$\begin{aligned}1 - \bar{\pi}B(\bar{k} - 1; n, p_0) - (1 - \bar{\pi})B(\bar{k}; n, p_0)& \\ < 1 - \bar{\pi}P(\bar{k} - 1; np_0) - (1 - \bar{\pi})P(\bar{k}; np_0).&\end{aligned} \tag{4.5}$$

Since $k = 2, 3, \cdots$ here, the result does not cover tests for which $k = 1$ leads to rejection, with or without randomization.

These properties apply also to two-sided tests. If the nominal significance level is less than a given number, the nominal probability of rejection in each tail is less than that number, and we make the above deductions about $\lambda = np_0$.

Our conclusions can be simplified to the following rules:

RULE 1. *The actual significance level of a nonrandomized test of the parameter of a binomial distribution is less than the approximate significance level based on the Poisson distribution if the approximate significance level is less than or equal to* .26, *except for the test which accepts the null hypothesis for* 0 *success and rejects for every positive number of successes.*

RULE 2. *The actual significance level of a randomized test of the parameter of a binomial distribution is less than the approximate significance level based on the Poisson distribution if the approximate significance level is less than or equal to* .08, *except possibly for tests which accept the null hypothesis with some positive probability for* 0 *success, reject the null hypothesis with some positive probability for one success and always reject the null hypothesis for more than one success.*

Rule 2 could possibly be improved, because for any particular $\bar{\pi}$ (< 1), (4.5) may hold without $B(\bar{k} - 1; n, \lambda/n)$ decreasing in n. However, a study of conditions for every $\bar{\pi}$ would be very complicated; for example, lemma 2 of Samuels [2] shows that if $\bar{\pi} = \frac{1}{2}$, then $\frac{1}{2}B(k - 1; n, \lambda/n) + \frac{1}{2}B(k; n, \lambda/n)$ is decreasing if $\lambda \leq [k(k - 1)]^{1/2}(n - 1)/\{[k(k - 1)]^{1/2} + [(n - k)(n - k - 1)]^{1/2}\}$.

Rule 2 can be improved by omitting tests for which $\bar{k} = 1$ or 2. Then the Poisson probability exceeds the binomial probability if the Poisson probability is at least $P(0; 2) = e^{-2} \sim .1353$ from theorem 4.1; application of theorem 4.2 to randomized tests for $\bar{k} \geq 3$ gives the criterion of $1 - P(3; 2) = 1 - 19e^{-2}/3 \sim .1429$.

4.2. *Confidence limits.* The upper confidence limit $\bar{p}$ for the parameter p of a binomial distribution based on a sample of k successes in n independent trials at confidence level $1 - \epsilon$ is the solution in p of $B(k; n, p) = \epsilon$. The upper confidence limit $\bar{\lambda}$ for the parameter λ of a Poisson distribution based on k occurrences at confidence level $1 - \epsilon$ is the solution in λ of $P(k; \lambda) = \epsilon$. An approximation to $\bar{p}$ is $\bar{\lambda}/n$. If $k + 1 \leq \bar{\lambda}$, then $B(k; n, \bar{\lambda}/n) \leq P(k, \bar{\lambda})$; since $B(k; n, p)$ decreases in p $[dB(k; n, p)/dp = -b(k; n - 1, p)]$, $\bar{p} \leq \bar{\lambda}/n$. Thus $\bar{\lambda}/n$ is a conservative upper confidence limit for p in the sense that the actual confidence level $1 - B(k; n, \bar{\lambda}/n)$ is greater than the nominal confidence level $1 - P(k; \bar{\lambda}) = 1 - \epsilon$. This is true if $1 - \epsilon \geq 1 - e^{-1}$ (approximately .6321). In practice, if $\bar{\lambda}/n > 1$, the upper limit for p is taken as one.

The lower confidence limit $\underline{p}$ for the parameter p of a binomial distribution based on k successes in n trials at confidence level $1 - \delta$ is the solution in p of $1 - B(k - 1; n, p) = \delta$. The lower confidence limit $\underline{\lambda}$ of a Poisson distribution based on k occurrences at confidence level $1 - \delta$ is the solution in λ of $1 - P(k - 1; \lambda) = \delta$. Then $\underline{\lambda}/n$ is an approximation to $\underline{p}$. If $\underline{\lambda} \leq k - 1$, then $P(k - 1; \underline{\lambda}) < B(k - 1; n, \underline{\lambda}/n)$ and $\underline{\lambda}/n < \underline{p}$. Thus $\underline{\lambda}/n$ is conservative since $B(k - 1; n, \underline{\lambda}/n) > P(k - 1; \underline{\lambda}) = 1 - \delta$. This is true if $1 - \delta \geq 2e^{-1}$ (approximately .7358). The conservative procedure then is $\underline{p} = 0$ if $k = 0$, $\underline{p} = 1 - (1 - \delta)^{1/n}$ if $k = 1$ (the solution of $B(0; n, p) = (1 - p)^n = 1 - \delta$), and $\underline{\lambda}/n$ if $k > 1$.

A confidence interval of confidence $1 - \epsilon - \delta$ is $(\underline{p}, \overline{p})$, where $\underline{p}$ and $\overline{p}$ are defined above. An approximate procedure is to use $(\underline{\lambda}/n, \overline{\lambda}/n)$, except that $\underline{\lambda}/n$ is replaced by $\underline{p}$ if $k = 1$, and $\overline{\lambda}/n$ is replaced by 1 if $\overline{\lambda}/n > 1$.

RULE 3. *If the confidence level of the Poisson approximate confidence limits (with $\underline{p}$ if $k = 1$) is at least .74, the actual confidence level is greater than the nominal level.*

5. An alternative approach to inequalities of sections 2 and 3

The following theorem supplements Hoeffding's Theorem stated in section 2.1 and yields a corollary which is stronger than theorem 2.1 but slightly weaker than theorem 2.3.

THEOREM 5.1. *If $F(k)$ and λ are defined as in section 2.1 with*

$$\lambda > [(n-1)/(n-3)][k - 1/(n-k)], \tag{5.1}$$

and if $p_i \leq \lambda/(n-1)$, $i = 1, \cdots, n$, then $F(k) \leq B(k; n, \lambda/n)$.

PROOF. The set of all vectors $(p_1, \cdots, p_n)$ which satisfy the hypothesis is compact, and $F(k)$ is a continuous function of p; hence, the supremum of $F(k)$ is attained. Suppose that the p_i's are not all equal; let $p_1 = \min_{i=1,\cdots,n} p_i$ and $p_2 = \max_{i=1,\cdots,n} p_i$. Let $f^*(j)$ and $F^*(j)$ be, respectively, the probabilities of j successes and of not more than j successes in trials 3 through n. Then

$$\begin{aligned} F(k) &= p_1p_2F^*(k-2) + [p_1(1-p_2) + p_2(1-p_1)]F^*(k-1) \\ &\qquad + (1-p_1)(1-p_2)F^*(k) \\ &= p_1p_2[f^*(k) - f^*(k-1)] - (p_1+p_2)f^*(k) + F^*(k). \end{aligned} \tag{5.2}$$

Since p_1 and p_2 are each at most $\lambda/(n-1)$, the sum of the remaining p_i's is at least $(n-3)\lambda/(n-1)$. Theorem 2 of [2] states that, if the sum of these p_i's is greater than $k - 1/(n-k)$, then $f^*(k) > f^*(k-1)$. Hence, if

$$\lambda > [(n-1)/(n-3)][k - 1/(n-k)], \tag{5.3}$$

we can increase $F(k)$ by replacing p_1 and p_2 by $(p_1 + p_2)/2$. Thus, under the hypothesis, the supremum of $F(k)$ is attained only when the p_i's are all equal, which gives the desired result.

If we take $p_1 = 0$, $p_2 = \cdots = p_n = \lambda/(n-1)$, we have the following corollary.

COROLLARY 5.1. *If $k < \lambda$, then $B(k; n, \lambda/n) > B(k; n-1, \lambda/(n-1))$ for all n sufficiently large so that*

$$\lambda > [(n-1)/(n-3)][k - 1/(n-k)]. \tag{5.4}$$

Note that the right-hand side of (5.4) is greater than the right-hand side of (2.14).

It is possible to obtain a result almost as good as theorem 3.1 by a method analogous to that used in proving theorem 2.1. We begin with a theorem of Hoeffding [1] which is more general than that in section 2.1.

GENERAL THEOREM OF HOEFFDING. *Let $g(k)$ be any function of the number of successes k in n independent trials, and let λ be a number between 0 and n. Then the maximum and minimum values of Eg among all choices of $p_1, \cdots, p_n$ with $p_1 + \cdots + p_n = \lambda$ are attained for choices of the following form: r of the p_i's are 0, s of the p_i's are one, and the remaining $n - r - s$ of the p_i's are equal to $(\lambda - s)/(n - r - s)$.*

If $g(j)$ is one for $j \leq k$ and 0 for $j > k$, then $Eg = F(k)$. Evaluation of Eg for the possible values of r and s shows that if $\lambda \leq k + 1$, then min Eg is not obtained with $s > 0$, and if $\lambda \leq k$, min Eg is not attained with $r > 0$. This gives half of (2.2), and the other half is attained similarly.

We now take $g(j)$ to be one if $j = k$, and 0 otherwise. Then $Eg = f(k)$. We prove the following theorem.

THEOREM 5.2. *If $\lambda \leq k - 1 - k^{1/2}$ or $\lambda \geq k + 1 + k^{1/2}$, then*

$$\max_{p_1 + \cdots + p_n = \lambda} f(k) = b(k; n, \lambda/n). \tag{5.5}$$

PROOF. From the general theorem of Hoeffding, we need only consider those choices of $p_1, \cdots, p_n$ with r of the p_i's equal to 0, s of them equal to one and the remaining $n - r - s$ equal to $(\lambda - s)/(n - r - s)$. Let us call such choices "candidates."

We shall show that if λ satisfies the hypothesis, and if $r > 0$ or $s > 0$, then there is another choice of the p_i's satisfying the constraint for which the probability of k successes is greater. To do this, we first note that

$$\begin{aligned} f(k) = {} & p_1p_2[f^*(k) - 2f^*(k-1) + f^*(k-2)] \\ & + (p_1 + p_2)[f^*(k) - f^*(k-1)] + f^*(k), \end{aligned} \tag{5.6}$$

where p_1 and p_2 are the probabilities of success on any two specified trials, and $f^*(k)$ is the probability of k successes in the remaining $n - 2$ trials. If $p_1 < p_2$ and the coefficient of p_1p_2 is positive, then we can increase $f(k)$ without altering the sum $p_1 + p_2$ by replacing p_1 and p_2 by $(p_1 + p_2)/2$.

It can be shown that

$$b(k; n, p) - 2b(k-1; n, p) + b(k-2; n, p) \tag{5.7}$$

is negative if and only if

$$\begin{aligned} k - [k(n + 2 - k)/(n + 1)]^{1/2} & \leq (n + 2)p \\ & \leq k + [k(n + 2 - k)/(n + 1)]^{1/2}, \end{aligned} \tag{5.8}$$

and hence is positive if

$$p \leq (k - 1 - k^{1/2})/(n + 1) \quad \text{or} \quad p \geq (k + k^{1/2})/(n + 1). \tag{5.9}$$

For a candidate with $r > 0$, we take $p_1 = 0$, $p_2 = (\lambda - s)/(n - r - s)$, while for a candidate with $s > 0$, we take $p_1 = (\lambda - s)/(n - r - s)$, $p_2 = 1$. Then the coefficient of p_1p_2 is

$$
(5.10)\quad \begin{aligned}
&b(k - s; n - r - s - 1, (\lambda - s)/(n - r - s)) \\
&\quad - 2b(k - s - 1; n - r - s - 1, (\lambda - s)/(n - r - s)) \\
&\quad + b(k - s - 2; n - r - s - 1, (\lambda - s)/(n - r - s)), \qquad \text{if} \quad r > 0; \\
&b(k - s + 1; n - r - s - 1, (\lambda - s)/(n - r - s)) \\
&\quad - 2b(k - s; n - r - s - 1, (\lambda - s)/(n - r - s)) \\
&\quad + b(k - s - 1; n - r - s - 1, (\lambda - s)/(n - r - s)), \qquad \text{if} \quad s > 0.
\end{aligned}
$$

From (5.9), the coefficient is positive and the theorem is proved.

Taking $r = 1$, $s = 0$, we have the following corollary.

COROLLARY 5.2. *If* $\lambda \leq k - 1 - k^{1/2}$ *or* $\lambda \geq k + 1 + k^{1/2}$, *then* $b(k; n, \lambda/n) \geq b(k - 1; n - 1, \lambda/(n - 1))$.

This, however, is weaker than theorem 3.1.

REFERENCES

[1] W. HOEFFDING, "On the distribution of the number of successes in independent trials," *Ann. Math. Statist.*, Vol. 27 (1956), pp. 713–721.

[2] STEPHEN M. SAMUELS, "On the number of successes in independent trials," *Ann. Math. Statist.*, Vol. 36 (1965), pp. 1272–1278.