Statistica Sinica 3(1993), 295-312

ON THE NUMBER OF SUCCESSES IN INDEPENDENT TRIALS

Y. H. Wang

Concordia University

Abstract: A unified combinatorial approach is used to obtain many theorems about S_n, the number of successes in n independent non-identical Bernoulli trials. The following results are, in particular, proved: (1) The variance of S_n increases as the set of success probabilities $\{p_i\}$ tends to be more and more homogeneous and attains its maximum as they become identical; (2) The density of S_n is unimodal: first increasing then decreasing; (3) Four different versions of Poisson's theorem; (4) An upper bound for the total variation between the distribution of S_n and that of the Poisson.

Key words and phrases: Bernoulli trials, Poisson's binomial distribution, unimodality, Poisson's theorem, total variation.

1. Introduction

Let S_n denote the total number of successes in n independent Bernoulli trials. If the trials are identical, the distribution of S_n and many of its properties and theorems are well known and well treated in many books on statistics and probability.

Historically, the distribution of S_n plays a very important role in the development of probability theory. It is for S_n that the very first forms of the law of large numbers (by Jacob Bernoulli in 1713), the central limit theorem (by Abraham De Moivre in 1714) and Poisson's theorem — the binomial-to-Poisson convergence theorem (by Simon Denis Poisson in 1837) were obtained.

If the trials are not identical, the probability of success at each trial varies from one trial to another, the distribution of S_n and many of its properties and theorems are not as well known, nor as readily available.

In this paper, we shall use a unified combinatorial approach to investigate the distribution of S_n for the non-identical Bernoulli model and systematically examine many of its theorems and properties. Even though most of the results in this paper are known, the derivations and the proofs we present here are elementary and unique.

The presentation of derivations in mathematics using combinatorial approach is often bogged down by messy notations. The unified approach we use in this paper was inspired by the notations introduced recently in Brown and Rinott (1988) in describing multivariate infinitely divisible distribution.

In recent times, Hoeffding (1956) was the first person to systematically look into the non-identical Bernoulli model and obtain many interesting results. His works have been followed and extended by many others. The most recent one is Nedelman and Wallenius (1986).

2. Preliminaries

Let n be a positive integer and $x = 0, 1, \ldots, n$. Define

$$\begin{cases} \mathcal{F}_x = \{A : A \subseteq \{1, \ldots, n\}, \quad |A| = x\} \\ \mathcal{P}_x = \{\pi(A) : A \in \mathcal{F}_x, \pi \text{ is a permutation in } \mathbf{R}^x\}, \end{cases} \tag{1}$$

where $|A|$ denotes the number of elements of A. If $A \in \mathcal{F}_x$, then $A = (i_1, \ldots, i_x)$ is an *ordered set* such that $i_j < i_k$ if $j < k$, and $\pi(A) = (\pi(i_1), \ldots, \pi(i_x))$ is a permutation of the elements of A.

There is only one element in $\mathcal{F}_n$, namely the n-tuple $(1, \ldots, n)$; but $|\mathcal{F}_x| = \binom{n}{x}$ is the number of subsets of size x of $\{1, \ldots, n\}$ and $|\mathcal{P}_n| = n!$ is the number of permutations of $\{1, \ldots, n\}$. For each fixed $A \in \mathcal{F}_x$, there are $x!$ corresponding elements in $\mathcal{P}_x$ which is the number of permutations of the elements of A, hence $|\mathcal{P}_x| = n!/(n - x)!$ equals $|\mathcal{P}_n|$ divided by $(n - x)!$.

The relations between $\mathcal{F}_n$, $\mathcal{P}_n$, $\mathcal{F}_x$ and $\mathcal{P}_x$ can be summarized as:

$$\begin{tikzcd} \mathcal{F}_n \arrow[r, "n!"] \arrow[d, swap, "\binom{n}{x}"] & \mathcal{P}_n \arrow[d, "\frac{1}{(n-x)!}"] \\ \mathcal{F}_x \arrow[r, "x!"] & \mathcal{P}_x \end{tikzcd}$$

Let α_i, β_i, $i = 1, \ldots, n$ be $2n$ non-zero real numbers. Then

$$\prod_{i=1}^{n} (\alpha_i + \beta_i) = \sum_{x=0}^{n} \sum_{A \in \mathcal{F}_x} \left\{ \left(\prod_{i \in A} \alpha_i \right) \left(\sum_{j \in A^c} \beta_j \right) \right\}, \tag{2}$$

where A^c denotes the complement of A and we take $\prod_{i \in A} \alpha_i = 1$ for $A \in \mathcal{F}_0 = \phi$.

An alternate expression for (2) is

$$\prod_{i=1}^{n} (\alpha_i + \beta_i) = \prod_{j=1}^{n} \beta_j \left[\sum_{x=0}^{n} \sum_{A \in \mathcal{F}_x} \left\{ \prod_{i \in A} (\alpha_i / \beta_i) \right\} \right]. \tag{2a}$$

Because $\prod_{i\in A} \alpha_i$ are *permutation invariant*, we have the following identity,

$$\sum_{A\in\mathcal{F}_x}\prod_{i\in A}\alpha_i = \Big(\sum_{A\in\mathcal{P}_x}\prod_{i\in A}\alpha_i\Big)\Big/x!, \quad \text{for all } x = 0,1,\ldots,n. \tag{3}$$

Looking at the sum $\sum_{A\in\mathcal{F}_x}\prod_{i\in A}\alpha_i$ as x iterated summations we have:

$$\begin{aligned}\sum_{A\in\mathcal{F}_x}\prod_{i\in A}\alpha_i &= \sum_{i_1=1}^{n-x+1}\alpha_{i_1}\left(\sum_{i_2=i_1+1}^{n-x}\alpha_{i_2}\cdots\Big(\sum_{i_x=i_{x-1}+1}^{n}\alpha_{i_x}\Big)\cdots\right)\\ &= \sum_{i_1=1}^{n}\alpha_{i_1}\left(\sum_{i_2\neq i_1}^{n}\alpha_{i_2}\cdots\left(\sum_{\substack{i_x\neq i_j\\ j=1,\ldots,x-1}}^{n}\alpha_{i_x}\right)\cdots\right)\Big/x!.\end{aligned} \tag{4}$$

For two integers x and y with $0\le x\le y\le n$, we can write

$$\Big(\sum_{A\in\mathcal{F}_x}\prod_{i\in A}\alpha_i\Big)\Big(\sum_{B\in\mathcal{F}_y}\prod_{i\in B}\alpha_i\Big) = \sum_{k=y}^{M}\sum_{C\in\mathcal{F}_k}\frac{(x+y-2r)!}{(x-r)!(y-r)!}\prod_{i\in C}\alpha_i^{s_i}, \tag{5}$$

where $s_i = 1$ or 2, $\sum_{i\in C} s_i = x+y$, $M = \min(x+y,n)$ and $r = x+y-k$.

Equality (5) above can be proved as follows: For $k = y, y+1,\ldots,M$ and $C\in\mathcal{F}_k$ fixed, in the expansion of $(\sum_{A\in\mathcal{F}_x}\prod_{i\in A}\alpha_i)(\sum_{B\in\mathcal{F}_y}\prod_{i\in B}\alpha_i)$, there are exactly $r = x+y-k$ of the s_i's equal to 2, and $k-r$ of the s_i's equal to 1, so that $2r+(k-r) = r+k = x+y$. The number of such occurrences is $\frac{[(x-r)+(y-r)]!}{(x-r)!(y-r)!}$ which is the number of ways of distributing $[(x-r)+(y-r)]$ distinguishable α's with $(x-r)$ in one group and $(y-r)$ in the other.

Denote $N = (1,\ldots,n)$ and define the cross product

$$N^x = N\times\cdots\times N$$

of N with itself x times. Then $\mathcal{P}_x$ is a subset of N^x and can be written as

$$\mathcal{P}_x = \Big\{(i_1,\ldots,i_x)\in N^x : i_j\neq i_k \text{ if } j\neq k\Big\}.$$

Denote by $\mathcal{B}_x$ the complement of $\mathcal{P}_x$ with N^x as the universal set. Then the multinomial expansion can be written as;

$$\left(\sum_{i=1}^{n}\alpha_i\right)^x = \sum_{A\in\mathcal{P}_x}\prod_{i\in A}\alpha_i + \sum_{A\in\mathcal{B}_x}\prod_{i\in A}\alpha_i. \tag{6}$$

We note here that for $x = 4$ the last term on the right-hand-side of (6) in the more familiar and conventional expression is

$$\sum_{A\in\mathcal{B}_4}\prod_{i\in A}\alpha_i = \sum_{i=1}^{n}\alpha_i^2\sum_{j,k\neq i}\alpha_j\alpha_k + \sum_{i=1}^{n}\alpha_i^3\sum_{j\neq i}\alpha_j + \sum_{i=1}^{n}\alpha_i^4.$$

3. Poisson's Binomial Distribution

We shall call the distribution of S_n in the independent but non-identical Bernoulli model "Poisson's binomial distribution". Historically Poisson (1837) was supposedly the first person to consider this extension of the binomial distribution. (See Cramér (1946) and Edwards (1960).) "Poisson's binomial distribution" is called "the Poisson-binomial distribution" by many authors, i.e. Le Cam (1960), Hodges and Le Cam (1960), Edwards (1960) and Chen (1974). We have always felt the use of the latter terminology in this context inappropriate, because it could be confused easily with (a) mixtures of binomial distributions and (b) compound Poisson distribution with binomial compounding distribution. (See Johnson and Kotz (1969, Pages 78 and 190). Our terminology " Poisson's binomial" was suggested by an anonymous member of the Editorial Board of *Statistica Sinica*, for which we express our gratitude.)

Denote by p_i the probability of success at the ith trial. Without loss of generality, we shall assume $0 < p_i < 1$ for all i. The extensions to $0 \leq p_i \leq 1$ are immediate, even though cumbersome reformulations of our statements may sometimes be required. Denote

$$\boldsymbol{p} = (p_1,\ldots,p_n), \quad I = (1,\ldots,1) \text{ and } \bar{\boldsymbol{p}} = \bar{p}I,$$

where $\bar{p} = \sum_{i=1}^{n} p_i/n$.

Define

$$f_n(x;\boldsymbol{p}) = \sum_{A\in\mathcal{F}_x}\Big(\prod_{i\in A}p_i\Big)\Big(\prod_{j\in A^c}(1-p_j)\Big). \tag{7}$$

For simplicity, we shall write f, $f(x)$, $f_n(x)$ or $f(x;\boldsymbol{p})$ for $f_n(x;\boldsymbol{p})$ and restore it whenever confusion may arise.

If $p_i = p$ for all i, (7) reduces to the usual binomial distribution

$$b(x) = \binom{n}{x}p^x(1-p)^{n-x}. \tag{8}$$

A useful property of f is that it is invariant under the permutation of $(p_1,\ldots,p_n)$. As in (2), an alternate expression for f is

$$f_n(x;\boldsymbol{p}) = \prod_{j=1}^{n}(1-p_j)\left[\sum_{A\in\mathcal{F}_x}\Big(\prod_{i\in A}\Big(\frac{p_i}{1-p_i}\Big)\Big)\right]. \tag{7a}$$

(Using different notations, the expression (7a) is used in Barbour, Holst and Janson (1992, Equation (1.4)).)

Combining (3) and (7a), a third expression for f is

$$f_n(x;\boldsymbol{p}) = \prod_{j=1}^{n}(1-p_j)\left[\sum_{A\in\mathcal{P}_x}\left(\prod_{i\in A}\left(\frac{p_i}{1-p_i}\right)\right)\right]\Big/x!. \tag{7b}$$

In this paper, we take the distribution of S_n to be "as is" in (7). For all the computations concerning S_n we use only (7) and its two alternates. A recourse taken by almost every author on this topic is to use the fact that $S_n = X_1 + \cdots + X_n$ is the sum of n independent Bernoulli random variables. We shall forgo using this fact for it does not make our job easier.

Summing both sides of (7) over $x = 0, 1, \ldots, n$ and using (2) we have $\sum_{x=0}^{n} f(x) = \prod_{i=1}^{n}(p_i + (1-p_i)) = 1$, for all n and $\boldsymbol{p}$. Thus (7) is a bona-fide *probability mass function* (p.m.f.) satisfying the following relation

$$f(x;\boldsymbol{p}) = f(n-x; I-\boldsymbol{p}), \qquad \text{for all } n,\ x \text{ and } \boldsymbol{p}. \tag{9}$$

Multiplying both sides of (7) by e^{tx}, summing over $x = 0, 1, \ldots, n$ and using (2), we obtain *the moment generating function* (m.g.f.) of S_n as

$$m(t) = \prod_{i=1}^{n}(1 - p_i + p_i e^t), \qquad \text{for all real } t. \tag{10}$$

Denote $p^{i_1\cdots i_k}$ as the $(n-k)$-tuple obtained from $\boldsymbol{p}$ by omitting $p_{i_1}, \ldots, p_{i_k}$, $k = 1, \ldots, n-1$.

Lemma 1. *For $1 \le k \le x < n$, Poisson's binomial* p.m.f. f_n *satisfies*

$$\begin{aligned} &x(x-1)\cdots(x-k+1)f_n(x;\boldsymbol{p}) \\ &= \sum_{i_1=1}^{n} p_{i_1}\left[\sum_{i_2\ne i_1}^{n} p_{i_2}\cdots\left(\sum_{\substack{i_k\ne i_j\\ j=1,\ldots,k-1}} p_{i_k}\left(f_{n-k}(x-k;\boldsymbol{p}^{i_1\cdots i_k})\right)\right)\cdots\right]. \end{aligned} \tag{11}$$

Proof. Using (7b) with $\alpha_i = p_i/(1-p_i)$, we have

$$\begin{aligned} &x(x-1)\cdots(x-k+1)f_n(x;\boldsymbol{p}) \\ &= \prod_{j=1}^{n}(1-p_j)\Big(\sum_{A\in\mathcal{P}_x}\prod_{i\in A}\alpha_i\Big)\Big/(x-k)! \end{aligned}$$

$$= \prod_{j=1}^{n}(1-p_j)\left[\sum_{i_1=1}^{n}\alpha_{i_1}\sum_{i_2\neq i_1}^{n}\alpha_{i_2}\cdots\sum_{\substack{i_k\neq i_j\\ j=1,\ldots,k-1}}\alpha_{i_k}\Big(\sum_{A\in\mathcal{P}_{x-k}}\prod_{i\in A}\alpha_i\Big)\Big/(x-k)!\right]$$

$$= \sum_{i_1=1}^{n}p_{i_1}\left(\sum_{i_2\neq i_1}^{n}p_{i_2}\cdots\left(\sum_{\substack{i_k\neq i_j\\ j=1,\ldots,k-1}}p_{i_k}f_{n-k}(x-k;\boldsymbol{p}^{i_1\ldots i_k})\right)\cdots\right).$$

Two special cases of (11) are worth mentioning. For $x = k$, it becomes

$$x!f_n(x;\boldsymbol{p}) = \sum_{i_1=1}^{n}p_{i_1}\left(\sum_{i_2\neq i_1}^{n}p_{i_2}\cdots\left(\sum_{\substack{i_x\neq i_j\\ j=1,\ldots,x-1}}p_{i_x}\left(f_{n-x}(0;\boldsymbol{p}^{i_1\ldots i_x})\right)\right)\cdots\right), \tag{12}$$

which is the same as (7b) in iterated summation form (4).

For $k = 1$, it reduces to the well-known and often-used identity

$$xf_n(x;\boldsymbol{p}) = \sum_{i=1}^{n}p_i f_{n-1}(x-1;\boldsymbol{p}^i), \quad \text{for all } x = 1,\ldots,n. \tag{13}$$

The identity (13) was first seen in the literature in Samuels (1965) who obtained it "by differentiating (10) with respect to e^t and identifying coefficients". It appeared later in Chen (1974) and again in many of his other publications. Chen derived it from an identity on conditional expectation due to Stein. (See Stein (1986).) It has also been used by many other authors mostly in problems concerning the Poisson approximation to Poisson's binomial distribution.

Here is another derivation of (13) which gives one a good feel of its meaning. Denote by E_i the event that the ith trial is a success. Then

$$\{S_n = x\} \subseteq \bigcup_{i=1}^{n}E_i, \quad \text{for all } x = 1,\ldots,n.$$

If a sample point ω is in $\{S_n = x\}$, then it is in exactly x of E_i's. Hence

$$xP(S_n = x) = \sum_{i=1}^{n}P(S_n = x|E_i)P(E_i),$$

which is another form of (13).

By (9), an equivalent expression for (13) is

$$(n-x)f_n(x;\boldsymbol{p}) = \sum_{i=1}^{n}(1-p_i)f_{n-1}(x;\boldsymbol{p}^i), \quad \text{for all } x = 0,\ldots,n-1. \tag{13a}$$

Summing both sides of (11) over $x = k, k+1, \ldots, n$, we obtain

$$E\Big[S_n(S_n-1)\cdots(S_n-k+1)\Big] = \sum_{i_1=1}^{n} p_{i_1}\left(\sum_{i_2\neq i_1}^{n} p_{i_2}\cdots\left(\sum_{\substack{i_k\neq i_j\\ j=1,\ldots,k-1}} p_{i_k}\right)\cdots\right). \quad (14)$$

It follows from (14) that

$$E(S_n) = \sum_{i=1}^{n} p_i \quad \text{and} \quad \mathrm{Var}(S_n) = \sum_{i=1}^{n} p_i(1-p_i). \quad (15)$$

About a century and a half ago, it was observed by Poisson (1837) (see also Edwards (1960)) that

$$\mathrm{Var}(S_n) = n\bar{p}(1-\bar{p}) - ns_p^2,$$

where $s_p^2 = \frac{1}{n}\sum_{i=1}^{n}(p_i-\bar{p})^2$ is the "variance" within $\{p_1, \ldots, p_n\}$. Therefore, *the variance of S_n increases as the set of probabilities $\{p_1, \ldots, p_n\}$ tends to be more and more homogeneous and attains its maximum as they become identical.*

We reformulate Poisson's observation as follows:

Theorem 1. *Denote by* $\mathrm{Var}(\boldsymbol{p})$ *the variance of* S_n *corresponding to the probability vector* $\boldsymbol{p} = (p_1, \ldots, p_n)$. *Then*

$$\mathrm{Var}(\boldsymbol{p}) \leq \mathrm{Var}(\boldsymbol{p}\Pi) \leq \mathrm{Var}(\bar{p}I). \quad (16)$$

for all $n \times n$ *doubly stochastic matrices* $\Pi = (\pi_{ij})$.

Proof. Let $g(p) = p(1-p)$ for $0 \leq p \leq 1$; then g is concave, satisfying

$$g(\alpha p_1 + (1-\alpha)p_2) \geq \alpha g(p_1) + (1-\alpha)g(p_2),$$

for all $0 \leq \alpha \leq 1$ and $0 \leq p_1, p_2 \leq 1$. Therefore, it follows from (15) that for all doubly stochastic matrices $\Pi = (\pi_{ij})$ and $\boldsymbol{p} = (p_1, \ldots, p_n)$,

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{p}\Pi) &= \sum_{j=1}^{n} g\Big(\sum_{i=1}^{n} p_i\pi_{ij}\Big) \\
&\geq \sum_{j=1}^{n}\sum_{i=1}^{n} \pi_{ij}g(p_i) \\
&= \sum_{i=1}^{n} g(p_i) = \mathrm{Var}(\boldsymbol{p}).
\end{aligned}$$

On the other hand, let Π be an arbitrary doubly stochastic matrix and denote $\boldsymbol{q} = \boldsymbol{p}\Pi$. Let $\Pi_0 = ((1/n))$ be an $n \times n$ square matrix with all the entries equal to $1/n$. Then $\boldsymbol{q}\Pi_0 = (\boldsymbol{p}\Pi)\Pi_0 = \boldsymbol{p}\Pi_0 = \bar{p}I$. (The first and the third equalities are evident while the second one can be easily verified.) Therefore, $\mathrm{Var}(\boldsymbol{p}\Pi) = \mathrm{Var}(\boldsymbol{q}) \leq \mathrm{Var}(\boldsymbol{q}\Pi_0) = \mathrm{Var}(\bar{p}I)$ for all doubly stochastic matrices Π.

An interpretation of Theorem 1 is as follows: Let $\bar{Y}_n = S_n/n$ be the "sample mean" of n independent, but not necessarily identical, Bernoulli trials. Then $E(\bar{Y}_n) = \bar{p}$ and $\mathrm{Var}(\bar{Y}_n) \leq \bar{p}(1-\bar{p})/n$, with equality holding if and only if $\boldsymbol{p} = \bar{p}I$. What this means in statistical terminology is that to estimate an unknown proportion $\bar{p}$, *the "unbiased sample mean" from a sequence of non-identically distributed Bernoulli random variables has smaller variance than the uniformly minimum variance unbiased estimate obtained by using the binomial density b with parameter $\bar{p}$.*

Here is a practical example: In a big American city, there are two candidates, one black and one white, running for the mayorship in an election. It is desired to estimate the voters' preference of the black candidate over the white candidate. Past experience indicates that the probability that a black voter prefers a black candidate is usually very high while for a white voter it is the opposite. Theorem 1 says that instead of taking a single sample of size n from the total population (simple random sampling) it is better to take two subsamples of sizes n_1 and n_2 (proportional to their respective subpopulations) with $n_1 + n_2 = n$ (stratified sampling). This of course is a well-established fact in sampling theory. (This example was suggested by Moishe Belinsky, for which the author is very grateful.)

Since $\mathrm{Var}(\bar{Y}_n) \leq \bar{p}(1-\bar{p})/n \leq 1/(4n)$, the next corollary follows immediately

Corollary 1. *For all $\varepsilon > 0$ and probability vector $\boldsymbol{p} = (p_1, \ldots, p_n)$,*

$$P\Big(|\bar{Y}_n - \bar{p}| > \varepsilon\Big) \leq 1/(4n\varepsilon^2).$$

It is well known that the binomial distribution b first increases monotonically, attains its maximum at $x = [(n+1)p]$ ($[s]$ denotes the integral part of s) and then decreases monotonically. (If $m = (n+1)p$ is an integer, $b(m) = b(m-1)$.) The corresponding property for Poisson's binomial distribution f was first noticed by Samuels (1965) who used an inequality of Newton to obtain

$$f^2(x) > f(x-1)f(x+1) \qquad \text{for } x = 1, \ldots, n-1, \tag{17}$$

(i.e. f is log concave). "Hence f is unimodal, first increasing, then decreasing, and the mode is either unique or shared by two adjacent integers", he observed. Inequality (17) was later quoted by many authors, but a comprehensive direct proof of it seems elusive.

Here we shall obtain an inequality sharper than (17) together with a simple proof.

Lemma 1. *Define*

$$g(x) = \sum_{A \in \mathcal{F}_x} \prod_{i \in A} \alpha_i, \quad \textit{for all real} \ \ \alpha_i > 0, \ i = 1, \ldots, n.$$

Then

$$g^2(x) > \left(\frac{x+1}{x}\right) g(x-1) g(x+1) \qquad \textit{for all} \ \ x = 1, \ldots, n-1. \tag{18}$$

Proof. We expand $g^2(x)$ and $g(x-1)g(x+1)$ according to (5). Since the first term in the expansion of $g^2(x)$ is strictly positive, we have

$$g^2(x) - \left(\frac{x+1}{x}\right) g(x-1)g(x+1) > \sum_{k=x+1}^{M} \sum_{C \in \mathcal{F}_k} \frac{[2(x-r)]!}{[(x-r)!]^2} K(x,r) \prod_{i \in C} \alpha_i^{s_i}, \tag{19}$$

where $s_i = 1$ or 2, $\sum_{i \in C} s_i = 2x$, $M = \min(2x, n)$, $r = 2x - k$ and

$$K(x,r) = \left[1 - \left(\frac{x+1}{x}\right) \frac{x-r}{x-r+1}\right].$$

For fixed $x = 1, \ldots, n-1$ $K(x,r) \geq 0$, for $r = 0, 1, \ldots$. Therefore the left-hand-side of (19) is greater than or equal to zero and the lemma follows.

Define

$$C(x) = \max\left(\frac{x+1}{x}, \frac{n-x+1}{n-x}\right), \quad \text{for all } x = 1, \ldots, n-1.$$

Theorem 2. *Poisson's binomial* p.m.f. *is unimodal, first increasing, then decreasing, and satisfies the following inequality*

$$f^2(x; \boldsymbol{p}) > C(x) f(x-1; \boldsymbol{p}) f(x+1; \boldsymbol{p}), \tag{20}$$

for all $\boldsymbol{p} = (p_1, \ldots, p_n)$ *and* $x = 1, \ldots, n-1$.

Proof. It follows from Lemma 1 and the alternate expression (7a) that

$$f^2(x; \boldsymbol{p}) > \left(\frac{x+1}{x}\right) f(x-1; \boldsymbol{p}) f(x+1; \boldsymbol{p}). \tag{21}$$

On the other hand, using (9) and Lemma 1 again we get

$$f^2(x; \boldsymbol{p}) > \left(\frac{n-x+1}{n-x}\right) f(x-1; \boldsymbol{p}) f(x+1; \boldsymbol{p}). \tag{22}$$

Both (21) and (22) are valid for $x = 1, \ldots, n-1$ and all $p = (p_1, \ldots, p_n)$.

Combining (21) and (22) we obtain (20).

To conclude this section, we present a numerical comparison of the binomial b with parameter $\bar{p}$ and Poisson's binomial f distributions, for $n = 5$, $p_1 = .10$, $p_2 = .30$, $p_3 = .45$, $p_4 = .60$, $p_5 = .80$ and $\bar{p} = .45$.

x	binomial	Poisson's binomial
0	.05033	.02772
1	.20589	.19010
2	.33691	.38690
3	.27565	.30150
4	.11277	.08730
5	.01845	.00648

As expected, both b and f are unimodal, first increasing, then decreasing and attain their unique maximum at $x = 2$. The values of f is more concentrated toward its mode at $x = 2$ than those of b. For $x = 2$ and 3, $f(x) > b(x)$, whereas $f(x) < b(x)$ for other values of x.

In the next section, most of the results concern the limiting behavior of S_n as n tends to infinity. Strictly speaking, the probability of success at the ith trial p_i should be written as p_{ni} to emphasize the dependence on n. But to save us from cumbersome notations, we drop the subscript n and believe that unnecessary confusions will not arise.

Before presenting the next section, we make some remarks here. In recent years, there are two popular approaches used by many authors working on problems in Poisson approximation: One is known as "the Stein-Chen method" in which functional equations such as (13) were derived by using conditional expectation and used to find upper bounds for the total variations such as (39) or to solve other problems in Poisson approximation. (See Chen (1974, 1975), Barbour and Hall (1984), Arratia, Goldstein and Gordon (1990) and the references cited there.) The Stein-Chen method is used *mainly* for sequences of weakly dependent (asymptotically independent) Bernoulli random variables with Poisson limit distribution. (See Steele (1990).) In many situations it is not the best method to use. For example to show that Poisson's binomial distribution is dominated by the Poisson distribution, the method we present in the next section (see Inequality (26)) is much simpler than the proof in Chen (1974) using the Stein-Chen method.

The other approach is called "the semi-group approach" which was first introduced into the problem of Poisson' theorem by Le Cam (1960). (See Pfeifer

(1985), Deheuvels and Pfeifer (1986) and the references cited there.) The semigroup approach is very much limited to sequences of *independent Bernoulli* random variables having *Poisson* as the limiting distribution.

To deal with sequences of random variables having strict dependent relation and/or with ranges in the set of integers and/or having limit distributions larger than Poisson, the above two approaches offer limited usefulness. Some possible recourses are offered in Wang (1986) and Lin and Wang (1993).

4. Poisson's Theorem

The Poisson p.m.f.

$$P(x) = e^{-\lambda}\lambda^x/x!, \quad x = 0, 1, 2, \ldots, \quad \lambda > 0,$$

was originally derived as a limit of the negative binomial distribution by Poisson (1837). But in the literature today it is generally known as the limit of the binomial distribution as $n \to \infty$ and $np \to \lambda$. The related theorem is called "Poisson's theorem".

Almost a century later, von Mises (1921) pointed out that the Poisson distribution is also the limit of Poisson's binomial p.m.f. provided that

$$\sum_{i=1}^{n} p_i = \lambda > 0 \text{ (fixed)} \quad \text{and} \quad \max_i p_i \to 0 \text{ as } n \to \infty. \tag{23}$$

His proof utilizes probability generating functions and can be found in Feller (1968). Later, Koopman (1950) showed that conditions (23) are also necessary.

In this section, we present several versions of Poisson's theorem for Poisson's binomial distribution. We first present two lemmas needed in the sequel. Lemma 3 follows immediately from the iterated summations (4) and shall be stated without proof.

Lemma 2. *If $\alpha_1, \ldots, \alpha_n$ are non-negative real numbers in the closed unit interval $[0, 1]$, then*

$$0 \le \exp\left(\sum_{i=1}^{n} -\alpha_i\right) - \prod_{i=1}^{n}(1-\alpha_i) \le \frac{1}{2}\sum_{i=1}^{n}\alpha_i^2, \quad \textit{for all} \quad n = 1, 2, \ldots. \tag{24}$$

Proof. For any $\alpha \ge 0$, we have

$$1 - \alpha \le e^{-\alpha} \le 1 - \alpha + \frac{1}{2}\alpha^2.$$

On the other hand, for $0 \le a_i,\ b_i \le 1$, $i = 1, \dots, n$, $c_1 = \prod_{j=2}^n b_j$, $c_i = (\prod_{j=1}^{i-1} a_j)(\prod_{j=i+1}^n b_j)$, we can write $\prod_{i=1}^n a_i - \prod_{i=1}^n b_i = \sum_{i=1}^n c_i(a_i - b_i)$. Therefore

$$0 \le \prod_{i=1}^n e^{-\alpha_i} - \prod_{i=1}^n (1-\alpha_i) \le \sum_{i=1}^n \left[e^{-\alpha_i} - (1-\alpha_i)\right] \le \frac{1}{2}\sum_{i=1}^n \alpha_i^2.$$

Lemma 3. *If $\lambda = \sum_{i=1}^n p_i$, then*

$$\sum_{A \in \mathcal{F}_x} \prod_{i \in A} p_i \le \lambda^x / x!, \quad x = 0, 1, 2, \dots, \tag{25}$$

with equality holding if and only if $x = 0$ or 1.

A consequence of Lemma 3 is that for $\lambda = \sum_{i=1}^n p_i$

$$f_n(x) \le \lambda^x/x! = Se^{-\lambda}\lambda^x/x! \qquad \text{for } x = 0, 1, 2, \dots, \tag{26}$$

where $S = e^{\lambda} > 0$. Therefore Poisson's binomial distribution is *dominated* by the Poisson distribution. As stated in Section 3 that Chen (1974) obtained similar result by using the Stein-Chen method. His proof was very long, complicated and hard to understand. Furthermore, he only showed the existence of S in (26) while here we specify the least possible value of S. Another more general method of deriving (26) can be found in Wang (1991).

Theorem 3 below is the first version of Poisson's theorem for Poisson's binomial distribution. Its sufficient part is more general than the result obtained by von Mises (1921). (Von Mises' two conditions are "$\sum_{i=1}^n p_i = \lambda$ and $\max_{1 \le i \le n} p_i \to 0$" which are equivalent to "$\sum_{i=1}^n p_i = \lambda$ and $\sum_{i=1}^n p_i^2 \to 0$" (as $n \to \infty$).) A much more general extension of Poisson's theorem in this direction is Theorem 3 in Wang (1989).

Theorem 3. *For a sequence of Poisson's binomial* p.m.f. $\{f_n\}$ *to have the following point-wise convergence*

$$\lim_{n \to \infty} f_n(x) = e^{-\lambda}\lambda^x/x!, \quad x = 0, 1, 2, \dots, \tag{27}$$

it is necessary and sufficient that the two limit conditions below hold:

$$\lambda_n = \sum_{i=1}^n p_i \to \lambda > 0 \quad \textit{and} \quad \sum_{i=1}^n p_i^2 \to 0 \quad (\textit{as } n \to \infty). \tag{28}$$

Proof. First we show that the two conditions in (28) are sufficient.

For $x = 0$, (27) follows from (24).

Let $\gamma_n = \max_{1\le i\le n} p_i$. Denote $\prod^{(n-t)} \alpha_i$ ($\sum^{(n-t)} \alpha_i$) as the product (the summation) of elements of a subset of size $(n-t)$ of $\{\alpha_1, \ldots, \alpha_n\}$. Then for $x = 1, 2, \ldots,$

$$\prod_{i=1}^{n}(1-p_i) \le \prod^{(n-x)}(1-p_i) \le \exp\Big(-\sum^{(n-x)} p_i\Big) \le e^{-\lambda_n + x\gamma_n} \tag{29}$$

and for $0 \le s \le x-1$,

$$\lambda_n \ge \sum^{(n-s)} p_i \ge \lambda_n - s\gamma_n. \tag{30}$$

By (11) with $k = x$, and (29) and (30), we obtain

$$f_n(0)\prod_{s=0}^{x-1}(\lambda_n - s\gamma_n) \le x! f_n(x) \le \lambda_n^x e^{-\lambda_n + x\gamma_n}. \tag{31}$$

Letting $n \to \infty$ and noting that

$$\gamma_n = \max_{1\le i\le n} p_i \le \Big(\sum_{i=1}^{n} p_i^2\Big)^{1/2},$$

(28) implies both ends of (31) converge to $\lambda^x e^{-\lambda}$ for all fixed x.

Next we show (28) also necessary for (27).

By (26) and the Lebesgue dominated convergence theorem, it follows that

$$\lim_{n\to\infty}\sum_{i=1}^{n} p_i = \lim_{n\to\infty} E(S_n) = \sum_{x=0}^{\infty} x e^{-\lambda}\lambda^x/x! = \lambda, \tag{32}$$

which proves the first part of (28). Similarly,

$$\lim_{n\to\infty}\sum_{i=1}^{n} p_i(1-p_i) = \lim_{n\to\infty} \mathrm{Var}(S_n) = \sum_{x=0}^{\infty}(x-\lambda)^2 e^{-\lambda}\lambda^x/x! = \lambda,$$

which, together with (32), concludes the second part of (28).

Theorem 4. *If*

$$\max_{1\le i\le n} p_i \to 0, \quad \textit{as} \;\; n \to \infty, \tag{33}$$

then a sequence of Poisson's binomial p.m.f. $\{f_n\}$ *converges to a non-degenerate* p.m.f. g *if and only if*

(a) g *is a Poisson* p.m.f. *with mean* λ,

(b) *both conditions in* (28) *hold.*

Proof. The "if" part is obvious. For the "only if" part, denote $S_n^i = S_n - X_i$, $i = 1, 2, \ldots, n$; then, by (33)

$$P(S_n \neq S_n^i) = p_i \to 0 \quad \text{(uniformly in } i \text{ as } n \to \infty). \tag{34}$$

In view of (33), we shall assume, without loss of generality, that $f_n(0; \boldsymbol{p}) > 0$ for all n. Since $f_{n-1}(x; \boldsymbol{p}^i)$ is the p.m.f. of S_n^i, it follows from (34) that

$$\lim_{n\to\infty} \frac{f_{n-1}(0; \boldsymbol{p}^i)}{f_n(0; \boldsymbol{p})} = 1 \quad \text{uniformly in } i. \tag{35}$$

We take $x = 1$ in (13) and divide both sides of it by $f_n(0; \boldsymbol{p})$; then (35) with $x = 1$ implies that

$$\lim_{n\to\infty} \sum_{i=1}^n p_i = \lim_{n\to\infty} \frac{\sum_{i=1}^n p_i f_{n-1}(0; \boldsymbol{p}^i)}{f_n(0; \boldsymbol{p})} = \lim_{n\to\infty} \frac{f_n(1; \boldsymbol{p}^i)}{f_n(0; \boldsymbol{p})} = \frac{g(1)}{g(0)} = \lambda, \quad \text{(say)}. \tag{36}$$

(The ratio $\lambda > 0$ because g is non-degenerate.)

Since $f_n(x)$ and $f_{n-1}(x)$ have the same limit $g(x)$ for all x, Equation (13) and the limit (36) imply that g must satisfy the functional equation

$$xg(x) = \lambda g(x-1) \quad \text{for all } x = 1, 2, \ldots. \tag{37}$$

It is well known that the solution of (37) is that g is Poisson with mean λ. (See Exercise #3.22 in Hogg and Craig (1978).)

Finally, by (33) and (36),

$$\sum_{i=1}^n p_i^2 \le \left(\max_i p_i\right)\left(\sum_{i=1}^n p_i\right) \to 0, \quad n \to \infty.$$

For two discrete p.m.f.'s g_1 and g_2 having a common domain in a countable set D, define a metric d by

$$d(g_1, g_2) = \sum_{k \in D} \left| g_1(k) - g_2(k) \right|.$$

The total variation d was first introduced into the problem of Poisson's theorem by Khintchine (1933). He obtained an upper bound for

$$d(b, P) = \sum_{x=0}^{\infty} \left| \binom{n}{x} p^x (1-p)^{n-x} - e^{-\lambda} \lambda^x / x! \right|, \tag{38}$$

and showed that it tends to zero as n tends to infinity and $np = \lambda$ (fixed). It was considered a big step from the pointwise convergence (27) (for b) to the convergence to zero of (38). In the next, we shall derive an upper bound for the total variation of Poisson's binomial and the Poisson p.m.f.. Our upper bound also tends to 0 under the two conditions in (28) with $\lambda = \sum_{i=1}^{n} p_i$.

Theorem 5. *For Poisson's binomial* p.m.f. $f_n(x)$

$$\sum_{x=0}^{\infty} \left| f_n(x) - e^{-\lambda}\lambda^x/x! \right| \leq 3e^{\lambda} \sum_{i=1}^{n} p_i^2 + e^{-\lambda} \frac{2\lambda^{n+1}}{n!(n+1-\lambda)}, \tag{39}$$

where $\lambda = \sum_{i=1}^{n} p_i$.

Proof. Let $B = \{x \in \{0,1,\ldots,n\}; e^{-\lambda}\lambda^x/x! > f_n(x)\}$ and

$$h(x) = \sum_{A \in \mathcal{F}_x} \prod_{i \in A} p_i, \quad x = 0,1,2,\ldots. \tag{40}$$

Then

$$\sum_{x=0}^{\infty} \left| f_n(x) - e^{-\lambda}\lambda/x! \right| = 2 \sum_{x \in B} \left(e^{-\lambda}\lambda^x/x! - f_n(x) \right) + 2 \sum_{x=n+1}^{\infty} e^{-\lambda}\lambda^x \big/ x!. \tag{41}$$

It follows from Lemma 3 and $1 - \alpha \leq e^{-\alpha}$ for all $\alpha \geq 0$ that

$$\prod_{i=1}^{n} (1-p_i) h(x) \leq \min\left(e^{-\lambda}\lambda^x/x!, f_n(x) \right), \quad x = 0,1,2,\ldots. \tag{42}$$

Using Inequality (42) twice, we can approximate the first term on the right-hand-side of (41) by

$$\begin{aligned}
&\leq 2 \sum_{x \in B} \left(e^{-\lambda}\lambda^x/x! - \prod_{i=1}^{n}(1-p_i)h(x) \right) \\
&\leq 2 \sum_{x=0}^{n} \left(e^{-\lambda}\lambda^x/x! - \prod_{i=1}^{n}(1-p_i)h(x) \right) \\
&= 2 \left[\sum_{x=0}^{n} \left(e^{-\lambda} - \prod_{i=1}^{n}(1-p_i) \right) \frac{\lambda^x}{x!} + \prod_{i=1}^{n}(1-p_i) \sum_{x=0}^{n} \left(\frac{\lambda^x}{x!} - h(x) \right) \right].
\end{aligned} \tag{43}$$

By Lemma 2 the first summation in (43) is bounded above by $\frac{1}{2}e^{\lambda}\sum_{i=1}^{n} p_i^2$. In the second summation, the first two terms vanish. For $x \geq 2$, let $\lambda_i = \lambda - p_i$ and $\mathcal{B}_x = N^x \backslash \mathcal{P}_x$ (as defined in Section 2); then we have

$$\left(\lambda^x/x! - h(x) \right) = \sum_{A \in \mathcal{B}_x} \prod_{i \in A} p_i/x! \leq \sum_{j=2}^{n} \binom{x}{j} \left[\sum_{i=1}^{n} p_i^j \lambda_i^{x-j} \right] \Big/ x!, \tag{44}$$

by equations (6) and (40). In (44), equality holds if and only if $x = 2$ and 3. By interchanging the order of summations, the right-hand-side of (44) can be further approximated by

$$\begin{aligned}
&(\lambda^x/x!)\Big[\sum_{i=1}^{n}\Big(1-(1-p_i/\lambda)^x - x(p_i/\lambda)(1-p_i/\lambda)^{x-1}\Big)\Big] \\
\le\ &(\lambda^x/x!)\Big[\sum_{i=1}^{n}\Big(1-(1-xp_i/\lambda) - x(p_i/\lambda)(1-(x-1)p_i/\lambda)\Big)\Big] \\
\le\ &\sum_{i=1}^{n} p_i^2\Big(\lambda^{x-2}\Big/(x-2)!\Big). \qquad (45)
\end{aligned}$$

Summing (44) over $x = 2, 3, \ldots$ and using (45), we get an upper bound for the second summation in (43) as $e^{\lambda}\sum_{i=1}^{n} p_i^2$.

The last term in (39) comes from the second term on the right-hand-side of (41)

$$\begin{aligned}
\sum_{x=n+1}^{\infty} e^{-\lambda}\lambda^x/x! \ &\le\ e^{-\lambda}\left\{\frac{\lambda^{n+1}}{(n+1)!}\left[1+\frac{\lambda}{(n+1)}+\left(\frac{\lambda}{(n+1)}\right)^2+\cdots\right]\right\} \\
&\le\ e^{-\lambda}\lambda^{n+1}/[n!(n+1-\lambda)].
\end{aligned}$$

This completes the proof.

It should be remarked here that the upper bound in (39) is not the sharpest possible bound. Our main purpose is to demonstrate that simple combinatorial approach works in this otherwise difficult case. Much sharper bounds can be found in Barbour and Hall (1984) for Poisson's binomial distribution and in Kennedy and Quine (1989) for the binomial distribution.

The following two corollaries are other versions of Poisson's theorem for Poisson's binomial distribution. Corollary 2 follows immediately from Theorem 5 or Corollary 3, while Corollary 3 is a consequence of Theorem 3 and the Lebesgue dominated convergence theorem.

Corollary 2. *For Poisson's binomial* p.m.f. $f_n(x)$, *if* $\lambda = \sum_{i=1}^{n} p_i$ *and* $\sum_{i=1}^{n} p_i^2 \to 0$ *as* $n \to \infty$, *then*

$$\lim_{n\to\infty}\sum_{i=1}^{\infty}\Big|f_n(x) - e^{-\lambda}\lambda^x/x!\Big| = 0.$$

Corollary 3. *For Poisson's binomial* p.m.f. f_n, *if* $\lambda = \sum_{i=1}^{n} p_i$ *and* $\sum_{i=1}^{n} p_i^2 \to 0$ *as* $n \to \infty$, *then*

$$\lim_{n\to\infty}\sum_{x=0}^{\infty} h(x)\Big|f_n(x) - e^{-\lambda}\lambda^x/x!\Big| = 0 \qquad (46)$$

for all non-negative function h with $\sum_{x=0}^{\infty} h(x)e^{-\lambda}\lambda^x \big/ x! < \infty$.

The mode of convergence (46) in Corollary 3 was first introduced into the problem of Poisson's theorem by Simons and Johnson (1971) for the binomial distribution. Their result was later shown to be true also for Poisson's distribution by Chen (1974). Corollary 3 is equivalent to the main theorem in Chen (1974). A more general result along the line can be found in Wang (1991). Recently, in Wang (1989, 1991 and 1992) and Wang and Ji (1993), we have obtained many results on this topic to cover the cases of compound Poisson and other related distributions. In Wang (1991) different modes of convergence are discussed and compared. According to a method used in that paper, the condition "$\lambda = \sum_{i=1}^n p_i$", in Corollaries 2 and 3, can be relaxed to "$\sum_{i=1}^n p_i \to \lambda > 0$".

References

Arratia, R., Goldstein, L. and Gordon, L. (1990). Poisson approximation and the Chen-Stein method. *Statist. Sci.* **5**, 403–434.

Barbour, A. D. and Hall, P. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Philios. Soc.* **95**, 473–480.

Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation.* Oxford University Press.

Brown, L. D. and Rinott, Y. (1988). Inequalities for multivariate infinitely divisible processes. *Ann. Probab.* **16**, 642–657.

Chen, L. H. Y. (1974). On the convergence of Poisson binomial to Poisson distributions. *Ann. Probab.* **2**, 178–180.

Chen, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3**, 534–545.

Cramér, H. (1946). *Mathmatical Methods of Statistics.* Princeton University Press, Princeton, N.J.

Deheuvels, P. and Pfeifer, D. (1986). A semigroup approach to Poisson approximation. *Ann. Probab.* **14**, 663–676.

Edwards, A. W. P. (1960). The meaning of binomial distribution. *Nature* **186**, 1074.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, **1**, 3rd edition. John Wiley, New York.

Hoeffding, W. (1956). On the distribution of the number of successes in independent trials. *Ann. Math. Statist.* **27**, 713–721.

Hodges, J. L., Jr. and Le Cam, L. (1960). The Poisson approximation to the Poisson binomial distribution. *Ann. Math. Statist.* **31**, 737-740.

Hogg, R. V. and Craig, A. T. (1978). *Introduction to Mathematical Statistics*, 4th edition. Macmillan Publishing Co., Inc, New York.

Johnson, N. L. and Kotz, S. (1969). *Distributions in Statistics-Discrete Distributions.* John Wiley.

Kennedy, J. E. and Quine, M. P. (1989). The total variation distance between the binomial and Poisson distributions. *Ann. Probab.* **17**, 396–400.

Khintchine, A. Ya. (1933). *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*. Springer-Verlag, Berlin.

Koopman, B. O. (1950). Necessary and sufficient conditions for Poisson's distribution. *Proc. Amer. Math. Soc.* **1**, 813–823.

Le Cam, L. (1960). An approximation theorem for the Poisson binomial distribution. *Pacific J. Math.* **10**, 1181–1197.

Lin, Z. Y. and Wang, Y. H. (1993). On the probability generating function of the sum of Markov random variables. (Submitted for publication.)

Nedelman, J. and Wallenius, T. (1986). Bernoulli trials, Poisson trials, surprising variances, and Jensen's inequality. *Amer. Statist.* **40**, 286–289.

Pfeifer, D. (1985). A semi-group setting for distance measures in connection with Poisson approximation. *Semigroup Forum.* **31**, 199–203.

Poisson, S. D. (1837). *Recherches sur la Probabilité des jugements en matière criminelle et en matière civile*. Bachelier, Paris.

Samuels, S. M. (1965). On the number of successes in independent trials. *Ann. Math. Statist.* **36**, 1272–1278.

Simons, G. and Johnson, N. L. (1971). On the convergence of binomial to Poisson distributions. *Ann. Math. Statist.* **42**, 1735–1736.

Steele, J. M. (1990). Comment in "Poisson approximation and the Chen-Stein method" by Arratia, R., Goldstein, L. and Gordon, L. *Statist. Sci.* **5**, 424–425.

Stein, C. M. (1986). *Approximate Computation of Expectations*. IMS, Hayward, California.

Von Mises, R. (1921). Über die Wahrscheinlichkeit seltener Ereignisse. *Z. angew. Math. Mech.* **1**, 121–124.

Wang, Y. H. (1986). Coupling methods in approximations. *Canad. J. Statist.* **14**, 69–74.

Wang, Y. H. (1989). From Poisson to compound Poisson approximations. *Math. Scientist.* **14**, 38–49.

Wang, Y. H. (1991). A compound Poisson convergence theorem. *Ann. Probab.* **19**, 452–455.

Wang, Y. H. (1992). Approximating the k-th order Markov Bernoulli chains. *J. Appl. Probab.* **29**, 861–868.

Wang, Y. H. and Ji, S. X. (1993). On the derivations of the compound Poisson distribution and process. *Statist. Probab. Lett.* (Accepted)

Department of Mathematics and Statistics, Concordia University, Montreal, PQ H3G 1M8, Canada.

(Received June 1991; accepted December 1992)