Statistical Science
2023, Vol. 38, No. 1, 108–119
https://doi.org/10.1214/22-STS852

The Poisson Binomial Distribution—Old & New[1]

Wenpin Tang and Fengmin Tang

Abstract. This is an expository article on the Poisson binomial distribution. We review lesser known results and recent progress on this topic, including geometry of polynomials and distribution learning. We also provide examples to illustrate the use of the Poisson binomial machinery. Some open questions of approximating rational fractions of the Poisson binomial are presented.

Key words and phrases: Distribution learning, geometry of polynomials, Poisson binomial distribution, Poisson/normal approximation, stochastic ordering, strong Rayleigh property.

1. INTRODUCTION

The binomial distribution is one of the earliest examples a college student encounters in his/her first course in probability. It is a discrete probability distribution of a sum of independent and identically distributed (i.i.d.) Bernoulli random variables, modeling the number of occurrence of some events in repeated trials. An integer-valued random variable X is called binomial with parameters (n, p), denoted as $X \sim \text{Bin}(n, p)$, if $\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, $0 \le k \le n$. It is well known that if n is large, the $\text{Bin}(n, p)$ distribution is approximated by the Poisson distribution for small p's, and is approximated by the normal distribution for larger values of p. See, for example, [63] for an educational tour.

Poisson [87] considered a more general model of independent trials, which allows heterogeneity among these trials. Precisely, an integer-valued random variable X is called Poisson binomial, and denoted as $X \sim \text{PB}(p_1, \dots, p_n)$ if

$$X \overset{(d)}{=} \xi_1 + \cdots + \xi_n,$$

where $\xi_1, \dots, \xi_n$ are independent Bernoulli random variables with parameters $p_1, \dots, p_n$. It is easily seen that the probability distribution of X is

$$\mathbb{P}(X = k) = \sum_{A \subset [n], |A| = k} \left(\prod_{i \in A} p_i \prod_{i \notin A} (1 - p_i) \right), \quad (1)$$

where the sum ranges over all subset of $[n] := \{1, \dots, n\}$ of size k.

The Poisson binomial distribution has a variety of applications such as reliability analysis [18, 59], survey sampling [29, 107], finance [40, 95], and engineering [44, 103]. Though this topic has been studied for a long time, the literature is scattered. For instance, the Poisson binomial distribution has different names in various contexts: Pólya frequency (PF) distribution, strong Rayleigh distribution, convolutions of heterogenous Bernoulli, etc. Researchers often work on some aspects of this subject, and ignore its connections to other fields. In late 1990s, Pitman [84] wrote a survey on the Poisson binomial distribution with focus on probabilizing combinatorial sequences. Due to its applications in modern technology (e.g., machine learning [25, 93], causal inference (Example 2)) and links to different mathematical fields (e.g., algebraic geometry, mathematical physics), we are motivated to survey recent studies on the Poisson binomial distribution. While most results in this paper are known in some form, several pieces are new (e.g., Section 4). The aim of this paper is to provide a guide to lesser known results and recent progress of the Poisson binomial distribution, mostly post 2000.

The rest of the paper is organized as follows. In Section 2, we review distributional properties of the Poisson binomial distribution. In Section 3, various approximations of the Poisson binomial distribution are presented. Section 4 is concerned with the Poisson binomial distribution and polynomials with nonnegative coefficients. There we discuss the problem of approximating rational fractions of Poisson binomial. Finally in Section 5, we consider some computational problems related to the Poisson binomial distribution.

Wenpin Tang is Assistant Professor, Department of Industrial Engineering and Operations Research, Columbia University, New York, New York 10027, USA (e-mail: wt2319@columbia.edu). Fengmin Tang is Graduate student, Institute for Computational & Mathematical Engineering, Stanford University, Stanford, California 94305, USA (e-mail: tfmin127@stanford.edu).

[1]In honor of Professor Jim Pitman for his 70th birthday

2. DISTRIBUTIONAL PROPERTIES OF POISSON BINOMIAL VARIABLES

In this section, we review a few distributional properties of the Poisson binomial distribution. For $X \sim \mathrm{PB}(p_1, \dots, p_n)$, we have

$$\mu := \mathbb{E}X = n\bar{p},$$
$$\sigma^2 := \operatorname{Var} X = n\bar{p}(1-\bar{p}) - \sum_{i=1}^n (p_i - \bar{p})^2, \tag{2}$$

where $\bar{p} := \sum_{i=1}^n p_i/n$. It is easily seen that by keeping $\mathbb{E}X$ (or $\bar{p}$) fixed, the variance of X is increasing as the set of probabilities $\{p_1, \dots, p_n\}$ gets more homogeneous, and is maximized as $p_1 = \cdots = p_n$. There is a simple interpretation in survey sampling: fixing the sample size from different communities (stratified sampling) provides better estimates of the overall probability than simple random sampling from the entire population.

The above observation motivates the study of stochastic orderings for the Poisson binomial distribution. The first result of this kind is due to Hoeffding [55], claiming that among all Poisson binomial distributions with a given mean, the binomial distribution is the most spread out.

THEOREM 2.1 ([55] Hoeffding's inequalities). *Let $X \sim \mathrm{PB}(p_1, \dots, p_n)$, and $\bar{X} \sim \mathrm{Bin}(n, \bar{p})$.*

1. *There are inequalities*

$$\mathbb{P}(X \le k) \le \mathbb{P}(\bar{X} \le k) \quad \textit{for } 0 \le k \le n\bar{p} - 1$$

and

$$\mathbb{P}(X \le k) \ge \mathbb{P}(\bar{X} \le k) \quad \textit{for } n\bar{p} \le k \le n.$$

2. *For any convex function $g : [n] \to \mathbb{R}$ in the sense that $g(k+2) - 2g(k+1) + g(k) > 0$, $0 \le k \le n-2$, we have*

$$\mathbb{E}g(X) \le \mathbb{E}g(\bar{X}),$$

where the equality holds if and only if $p_1 = \cdots = p_n = \bar{p}$.

The part (2) in Theorem 2.1 indicates that among all Poisson binomial distributions, the binomial is the largest one in convex order. The original proof of Theorem 2.1 was brute-force, and it was soon generalized by using the idea of majorization and Schur convexity, see Theorem 2.2(1). This result was also extended to the multidimensional setting [11], and to nonnegative random variables [10], Proposition 3.2. See also [76] for interpretations. Next, we give several applications of Hoeffding's inequalities.

EXAMPLE 1.

1. Monotonicity of binomials. Fix $\lambda > 0$. By taking $(p_1, \dots, p_n) = (0, \frac{\lambda}{n-1}, \dots, \frac{\lambda}{n-1})$, we get for $X \sim \mathrm{Bin}(n-1, \frac{\lambda}{n-1})$ and $X' \sim \mathrm{Bin}(n, \frac{\lambda}{n})$,

$$\mathbb{P}(X \le k) < \mathbb{P}(X' \le k) \quad \text{for } k \le \lambda - 1,$$

and

$$\mathbb{P}(X \le k) > \mathbb{P}(X' \le k) \quad \text{for } k \ge \lambda.$$

Similarly, by taking $(p_1, \dots, p_n) = (1, \frac{\lambda-1}{n-1}, \dots, \frac{\lambda-1}{n-1})$, we get for $X \sim \mathrm{Bin}(n-1, \frac{\lambda-1}{n-1})$ and $X' \sim \mathrm{Bin}(n, \frac{\lambda}{n})$,

$$\mathbb{P}(X \le k-1) < \mathbb{P}(X' \le k) \quad \text{for } k \le \lambda - 1,$$

and

$$\mathbb{P}(X \le k-1) > \mathbb{P}(X' \le k) \quad \text{for } k \ge \lambda.$$

These inequalities were used in [5] to derive the monotonicity of error in approximating the binomial distribution by a Poisson distribution. By letting $X \sim \mathrm{Bin}(n, p)$ and $Y \sim \mathrm{Poi}(np)$, they proved $\mathbb{P}(X \le k) - \mathbb{P}(Y \le k)$ is positive if $k \le n^2 p/(n+1)$ and is negative if $k \ge np$. The result quantifies the error of confidence levels in hypothesis testing when approximating the binomial distribution by a Poisson distribution.

2. Darroch's rule. It is well known that a Poisson binomial variable has either one, or two consecutive modes. By an argument in the proof of Hoeffding's inequalities, Darroch [32], Theorem 4, showed that the mode m of the Poisson binomial distribution differs from its mean μ by at most 1. Precisely, he proved that

$$m = \begin{cases} k & \text{if } k \le \mu < k + \dfrac{1}{k+2}, \\ k \text{ or } k+1 & \\ \quad \text{if } k + \dfrac{1}{k+2} \le \mu \le k + 1 - \dfrac{1}{n-k+1}, & \\ k+1 & \\ \quad \text{if } k + 1 - \dfrac{1}{n-k+1} < \mu \le k+1. & \end{cases} \tag{3}$$

This result was reproved in [94]. See also [62] for a similar result concerning the median.

3. Azuma–Hoeffding inequality. One of the most famous result of Hoeffding is the Azuma–Hoeffding inequality [7, 56]: for independent random variables $\xi_1, \dots, \xi_n$ with $0 \le \xi_i \le 1$,

$$\mathbb{P}\left(\sum_{i=1}^n \xi_i \ge t\right) \le \left(\frac{\mu}{t}\right)^t \left(\frac{n-\mu}{n-t}\right)^{n-t} \quad \text{for } t > \mu, \tag{4}$$

where $\mu := \sum_{i=1}^n \mathbb{E}\xi_i$. We show how to derive the deviation inequality (4) via Hoeffding's inequalities (Theorem 2.1). In fact, ξ_i is sub-Bernoulli [109] in the sense that its moments are all bounded by those of Bernoulli variables with the same mean. Thus, the moments of $\sum_{i=1}^n \xi_i$ are bounded by those of the Poisson binomial variable $\mathrm{PB}(\mathbb{E}\xi_1, \dots, \mathbb{E}\xi_n)$. By Theorem 2.1(2) with $g(x) = x^k$, $k \ge 1$, the moment generating function of $\sum_{i=1}^n \xi_i$ is bounded by that of the binomial variable $\mathrm{Bin}(n, \mu/n)$. It then suffices to apply the Chernoff bound to get the Azuma–Hoeffding inequality (4).

To proceed further, we need some vocabularies. Let $\{x_{(1)}, \dots, x_{(n)}\}$ be the order statistics of $\{x_1, \dots, x_n\}$.

DEFINITION 1. The vector $\boldsymbol{x}$ is said to majorize the vector $\boldsymbol{y}$, denoted as $\boldsymbol{x} \succeq \boldsymbol{y}$, if

$$\sum_{i=1}^k x_{(i)} \le \sum_{i=1}^k y_{(i)} \quad \text{for } k \le n-1 \quad \text{and}$$
$$\sum_{i=1}^n x_{(i)} = \sum_{i=1}^n y_{(i)}.$$

See [72] for background and development on the theory of majorization and its applications. The following theorem gives a few lesser known variants of Hoeffding's inequalities.

THEOREM 2.2. *Let* $X \sim \mathrm{PB}(p_1, \dots, p_n)$ *and* $X' \sim \mathrm{PB}(p'_1, \dots, p'_n)$, *and* $Y \sim \mathrm{Bin}(n, p)$.

1. [49, 107] *If* $(p_1, \dots, p_n) \succeq (p'_1, \dots, p'_n)$, *then*
$$\mathbb{P}(X \le k) \le \mathbb{P}(X' \le k) \quad \text{for } 0 \le k \le n\bar{p} - 2,$$
and
$$\mathbb{P}(X \le k) \ge \mathbb{P}(X' \le k) \quad \text{for } n\bar{p} + 2 \le k \le n.$$
Moreover, $\mathrm{Var}(X) \le \mathrm{Var}(X')$.
2. [86] *If* $(-\log p_1, \dots, -\log p_n) \succeq (-\log p'_1, \dots, -\log p'_n)$, *then* X *is stochastically larger than* X', *that is,* $\mathbb{P}(X \ge k) \le \mathbb{P}(X' \ge k)$ *for all* k.
3. [19] X *is stochastically larger than* Y *if and only if* $p \le (\prod_{i=1}^n p_i)^{\frac{1}{n}}$, *and* X *is stochastically smaller than* Y *if and only if* $p \ge 1 - (\prod_{i=1}^n (1-p_i))^{\frac{1}{n}}$. *Consequently, if* $(\prod_{i=1}^n p_i)^{\frac{1}{n}} \ge 1 - (\prod_{i=1}^n (1-p'_i))^{\frac{1}{n}}$ *then* X *is stochastically larger than* X'.

The proof of Theorem 2.2 relies on the fact that $\boldsymbol{x} \succeq \boldsymbol{y}$ implies the components of $\boldsymbol{x}$ are more spread out than those of $\boldsymbol{y}$. For example in part (1), it boils down to proving if $k \le n\bar{p} - 2$, $\mathbb{P}(X \le k)$ is a Schur concave function in $\boldsymbol{p}$, meaning its value increases as the components of $\boldsymbol{p}$ are less dispersed. Part (3) gives a sufficient condition of stochastic orderings for the Poisson binomial distribution. A simple necessary and sufficient condition remains open. See also [17, 18, 20, 54, 96, 108] for further results.

3. APPROXIMATION OF POISSON BINOMIAL DISTRIBUTIONS

In this section, we discuss various approximations of the Poisson binomial distribution. Pitman [84], Section 2, gave an excellent survey on this topic in the mid-1990s. We complement the discussion with recent developments. In the sequel, $\mathcal{L}(X)$ denotes the distribution of a random variable X.

Poisson approximation. Le Cam [70] gave the first error bound for Poisson approximation of the Poisson binomial distribution. The following theorem is an improvement of Le Cam's bound.

THEOREM 3.1 ([8]). *Let* $X \sim \mathrm{PB}(p_1, \dots, p_n)$ *and* $\mu := \sum_{i=1}^n p_i$. *Then*

$$\frac{1}{32}\min\left(1, \frac{1}{\mu}\right)\sum_{i=1}^n p_i^2 \le d_{\mathrm{TV}}(\mathcal{L}(X), \mathrm{Poi}(\mu)) \le \frac{1-e^{-\mu}}{\mu}\sum_{i=1}^n p_i^2, \tag{5}$$

where $d_{\mathrm{TV}}(\cdot,\cdot)$ *is the total variation distance.*

The proof of Theorem 3.1 relies on the Stein–Chen identity: by writing $X = \sum_{i=1}^n \xi_i$ with $\xi_1, \dots, \xi_n$ independent Bernoulli random variables with parameters $p_1, \dots, p_n$,

$$\mathbb{E}(\mu f(X+1) - Xf(X)) = \sum_{i=1}^n p_i^2 \mathbb{E}(f(X - \xi_i + 2) - f(X - \xi_i + 1)),$$

where f is any real-valued function on the nonnegative integers. The inequalities (5) are then obtained by a suitable choice of f. It is easily seen from (5) that the Poisson approximation of the Poisson binomial is good if $\sum_{i=1}^n p_i^2 \ll \sum_{i=1}^n p_i$, or equivalently $\mu - \sigma^2 \ll \mu$. There are two cases:

- For small μ, the upper bound in (5) is sharp. In particular, for $\mu \le 1$, by taking $p_1 = \mu$ and $p_2 = \dots = p_n = 0$, we have
$$d_{\mathrm{TV}}(\mathcal{L}(X), \mathrm{Poi}(\mu)) = \mu(1 - e^{\mu}) = \frac{1-e^{-\mu}}{\mu}\sum_{i=1}^n p_i^2.$$
- For large μ, the approximation error is of order $\sum_{i=1}^n p_i^2 / \sum_{i=1}^n p_i$.

As pointed out in [61], the constant $1/32$ in the lower bound can be improved to $1/14$. See [9] for a book-length treatment, and [90] for sharp bounds. A powerful tool to study the approximation of the sum of (possibly dependent) random variables is Stein's method of exchangeable pairs, see [26]. For instance, a simple proof of the upper bound in (5) was given in [26], Section 3.

The Poisson approximation can be viewed as a mean-matching procedure. The failure of the Poisson approximation is due to a lack of control in variance. A typical example is where all p_i's are bounded away from 0, so that μ is large and $\sum_{i=1}^n p_i^2 / \sum_{i=1}^n p_i$ is of constant order. To deal with these cases, Röllin [89] considered a mean/variance-matching procedure. To present further results, we need the following definition.

DEFINITION 2. An integer-valued random variable X is said to be translated Poisson distributed with parameters (μ, σ^2), denoted as $\mathrm{TP}(\mu, \sigma^2)$, if $X - \mu + \sigma^2 + \{\mu - \sigma^2\} \sim \mathrm{Poi}(\sigma^2 + \{\mu - \sigma^2\})$, where $\{\cdot\}$ is the fraction part of a positive number.

It is easy to see that a $\mathrm{TP}(\mu, \sigma^2)$ random variable has mean μ, and variance $\sigma^2 + \{\mu + \sigma^2\}$ which is between σ^2 and $\sigma^2 + 1$. The following theorem gives an upper bound in total variation between a Poisson binomial variable and its translated Poisson approximation.

THEOREM 3.2 ([89]). *Let* $X \sim \mathrm{PB}(p_1, \dots, p_n)$, *and* $\mu := \sum_{i=1}^n p_i$ *and* $\sigma^2 := \sum_{i=1}^n p_i(1-p_i)$. *Then*

$$d_{\mathrm{TV}}\big(\mathcal{L}(X), \mathrm{TP}(\mu, \sigma^2)\big) \le \frac{2 + \sqrt{\sum_{i=1}^n p_i^3(1-p_i)}}{\sigma^2}, \tag{6}$$

where $d_{\mathrm{TV}}(\cdot, \cdot)$ *is the total variation distance.*

Theorem 3.2 is a consequence of a more general result of Stein's exchangeable pairs for translated Poisson approximation. Note that if all p_i's are bounded away from 0 and 1, the approximation error is of order $1/\sqrt{n}$ which is optimal and is comparable to the normal approximation error (see Theorem 3.4(2)). See [77] for the most up-to-date results of the Poisson approximation. Now we give an application of translated Poisson approximation in observational studies.

EXAMPLE 2. Sensitivity analysis. In matched-pair observational studies, a sensitivity analysis assesses the sensitivity of results to hidden bias. Here we follow a modern approach of Rosenbaum [92], Chapter 4. More precisely, the sample consists of n matched pairs indexed by $k = 1, \dots, n$, and units in each pair are indexed by $i = 1, 2$. The pair k is matched on a set of observed covariates $\boldsymbol{x}_{k1} = \boldsymbol{x}_{k2}$, and only one unit in each pair receives the treatment. Let Z_{ki} be the treatment assignment, so $Z_{k1} + Z_{k2} = 1$. Common test statistics for matched pairs are sign-score statistics of the form: $T = \sum_{k=1}^n d_k(c_{k1}Z_{k1} + c_{k2}Z_{k2})$, where $d_k \ge 0$ and $c_{ki} \in \{0, 1\}$. Here c_{ki} represents the potential outcome which depends on the response $(r_{11}, r_{12}, \dots, r_{n1}, r_{n2})$. For instance, in the Wilcoxon's signed rank test: $c_{k1} = 1$ if $r_{k1} > r_{k2}$ and $c_{k1} = 0$ otherwise, and similarly, $c_{k2} = 1$ if $r_{k2} > r_{k1}$ and $c_{k2} = 0$ otherwise, so $c_{k1} = c_{k2} = 0$ if $r_{k1} = r_{k2}$. For simplicity, we take $d_k = 1$ and the statistics of interest are

$$T = \sum_{k=1}^n (c_{k1}Z_{k1} + c_{k2}Z_{k2}), \tag{7}$$

where $c_{k1}Z_{k1} + c_{k2}Z_{k2}$ is Bernoulli distributed with parameter $p_k := c_{k1}\pi_k + c_{k2}(1-\pi_k)$ with $\pi_k := \mathbb{P}(Z_{k1} = 1)$. So $T \sim \mathrm{PB}(p_1, \dots, p_n)$. For $1 \le k \le n$, let $\Gamma_k := \pi_k/(1-\pi_k)$, which equals to 1 if there is no hidden bias.

The goal is to make inference on T with different choices of $(\pi_1, \dots, \pi_n)$ and understand which choices explain away the conclusion we draw from the null hypothesis (i.e., there is no hidden bias). Thus, we are interested in the set

$$\mathcal{R}(t, \alpha) := \big\{(\pi_1, \dots, \pi_n) : \mathbb{P}(T \ge t) \le \alpha\big\},$$

on the boundary of which the conclusion assuming no hidden bias is turned over. However, direct computation of $\mathcal{R}(t, \alpha)$ seems hard. A routine way to solve this problem is to approximate $\mathcal{R}(t, \alpha)$ by a regular shape. To this end, we consider the following optimization problem:

$$\begin{aligned} &\max \Gamma, \\ &\text{s.t. } \max_{\boldsymbol{\pi} \in C_\Gamma} \mathbb{P}\big(T(\pi_1, \dots, \pi_n) \ge t\big) \le \alpha, \end{aligned} \tag{8}$$

where C_Γ is a constraint region. For instance, $C_\Gamma := \{\boldsymbol{\pi} : \frac{1}{1+\Gamma} \le \pi_k \le \frac{\Gamma}{1+\Gamma}\}$ corresponds to the worst-case sensitivity analysis. By the translated Poisson approximation, the quantity $\max_{\boldsymbol{\pi} \in C_\Gamma} \mathbb{P}(T(\pi_1, \dots, \pi_n) \ge t)$ can be evaluated by the following problem which is easy to solve:

$$\begin{aligned} &\min_{A \in \{0, \dots, K\}} \min_{\boldsymbol{\pi} \in C_\Gamma} \sum_{k=0}^K \frac{\lambda^k e^{-\lambda}}{k!} \\ &\text{s.t. } K = t - A, \lambda = \sum_{k=1}^n p_k - A, \\ &A \le \sum_{k=1}^n p_k^2 < A + 1. \end{aligned} \tag{9}$$

Normal approximation. The normal approximation of the Poisson binomial distribution follows from Lyapunov or Lindeberg central limit theorem, see, for example, [13], Section 27. Berry and Esseen independently discovered an error bound in terms of the cumulative distribution function for the normal approximation of the sum of independent random variables. Subsequent improvements were obtained by [79, 82, 97, 105] via Fourier analysis, and by [27, 28, 75, 104] via Stein's method.

Let $\phi(x) := \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$ be the probability density function of the standard normal distribution, and $\Phi(x) := \int_{-\infty}^x \phi(y)\, dy$ be its cumulative distribution function. The following theorem provides uniform bounds for the normal approximation of Poisson binomial variables.

THEOREM 3.3. *Let* $X \sim \mathrm{PB}(p_1, \dots, p_n)$, *and* $\mu := \sum_{i=1}^n p_n$ *and* $\sigma^2 := \sum_{i=1}^n p_i(1-p_i)$.

1. [85], *Theorem* 11.2, *There is a universal constant* $C > 0$ *such that*

$$\max_{0 \le k \le n} \left| \mathbb{P}(X = k) - \phi\left(\frac{k-\mu}{\sigma}\right) \right| \le \frac{C}{\sigma}. \tag{10}$$

2. [97] *We have*

$$\max_{0 \le k \le n} \left| \mathbb{P}(X \le k) - \Phi\left(\frac{k-\mu}{\sigma}\right) \right| \le \frac{0.7915}{\sigma}. \tag{11}$$

Other than uniform bounds (10)–(11), several authors [16, 50, 88] studied error bounds for the normal approximation in other metrics. For μ, ν two probability measures, consider:

- L^r metric

$$d_r(\mu,\nu) := \left(\int_{-\infty}^{\infty} |\mu(-\infty,x] - \nu(-\infty,x]|^r\, dx\right)^{\frac{1}{r}},$$

- Wasserstein's r metric

$$\mathcal{W}_r(\mu,\nu) := \inf_{\pi}\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x-y|^r \pi(dx\,dy)\right)^{\frac{1}{r}},$$

where the infimum runs over all probability measures π on $\mathbb{R}\times\mathbb{R}$ with marginals μ and ν.

Specializing these bounds to the Poisson binomial distribution, we get the following result.

THEOREM 3.4. *Let $X \sim \mathrm{PB}(p_1,\dots,p_n)$, and $\mu := \sum_{i=1}^n p_n$ and $\sigma^2 := \sum_{i=1}^n p_i(1-p_i)$.*

1. [83], *Chapter V, There exists a universal constant $C > 0$ such that*

$$d_r(\mathcal{L}(X), \mathcal{N}(\mu,\sigma^2)) \le \frac{C}{\sigma} \quad \text{for all } r \ge 1. \tag{12}$$

2. [16, 88] *For each $r \ge 1$, there exists a constant $C_r > 0$ such that*

$$\mathcal{W}_r(\mathcal{L}(X), \mathcal{N}(\mu,\sigma^2)) \le \frac{C_r}{\sigma}. \tag{13}$$

Goldstein [50] proved L^r bound (12) for $r = 1$ with $C = 1$ via zero bias transformation. The general case follows from the inequality $d_r(\mu,\nu)^r \le d_\infty(\mu,\nu)^{r-1} d_1(\mu,\nu)$ together with L^1 bound and the uniform bound (11). By the Kantorovich–Rubinstein duality, $d_1(\mu,\nu) = \mathcal{W}_1(\mu,\nu)$. So the bound (13) holds for $r = 1$ with $C_1 = 1$. For general r, the bound (13) is a consequence of the fact that for $Z = \sum_{i=1}^n \xi_i$ with ξ_i's independent, $\mathbb{E}\xi_i = 0$ and $\sum_{i=1}^n \mathrm{Var}(\xi_i) = 1$,

$$\mathcal{W}_r(\mathcal{L}(Z), \mathcal{N}(0,1)) \le C_r\left(\sum_{i=1}^n \mathbb{E}|Z_i|^{r+1}\right)^{\frac{1}{r}}.$$

This result was proved in [88] for $1 \le r \le 2$, and generalized to all $r \ge 1$ in [16].

Binomial approximation. The binomial approximation of the Poisson binomial is lesser known. The first result of this kind is due to Ehm [41] who proved that for $X \sim \mathrm{PB}(p_1,\dots,p_n)$,

$$\begin{aligned} &d_{\mathrm{TV}}(\mathcal{L}(X), \mathrm{Bin}(n,\mu/n)) \\ &\quad \le \frac{1-(\mu/n)^{n+1} - (1-\mu/n)^{n+1}}{(n+1)(1-\mu/n)\mu/n} \sum_{i=1}^n (p_i - \mu/n)^2. \end{aligned} \tag{14}$$

Elm's approach was extended to a Krawtchouk expansion in [91]. The advantage of the binomial approximation over the Poisson approximation is justified by the following result due to Choi and Xia [31].

THEOREM 3.5. *Let $X \sim \mathrm{PB}(p_1,\dots,p_n)$, and $\mu := \sum_{i=1}^n p_n$. For $m \ge 1$, let $d_m := d_{\mathrm{TV}}(\mathcal{L}(X), \mathrm{Bin}(m,\mu/m))$. Then for $m > \max\{\frac{\mu^2}{\lfloor\mu\rfloor - 1 - (1+\{\mu\})^2}, n\}$,*

$$d_m < d_{m+1} < \cdots < d_{\mathrm{TV}}(\mathcal{L}(X), \mathrm{Poi}(\mu)), \tag{15}$$

where $\lfloor\cdot\rfloor$ is the integer part and $\{\cdot\}$ is the fractional part of a positive number.

It is easily seen from Theorem 3.5 that for $\mu \ge 6$ and $n \ge \frac{(\lfloor\mu\rfloor+1)^2}{\lfloor\mu\rfloor-5}$, the $\mathrm{Bin}(n,\frac{\mu}{n})$ approximation is strictly better than the Poisson approximation. It was also conjectured that the best $\mathrm{Bin}(m,\frac{\mu}{m})$ approximation is achieved for $m = \lfloor\frac{\mu^2}{\sum_{i=1}^n p_i^2}\rfloor$ by a mean/variance matching argument. See also [9, 80] for multiparameter binomial approximations, and [98] for the Pólya approximation of the Poisson binomial distribution.

4. POISSON BINOMIAL DISTRIBUTIONS AND POLYNOMIALS WITH NONNEGATIVE COEFFICIENTS

In this section, we discuss aspects of the Poisson binomial distribution related to polynomials with nonnegative coefficients. For $X \sim \mathrm{PB}(p_1,\dots,p_n)$, the probability generating function (PGF) of X is

$$f(u) := \mathbb{E}u^X = \prod_{i=1}^n (p_i u + 1 - p_i). \tag{16}$$

It is easy to see that f is a polynomial with all nonnegative coefficients, and all of its roots are real negative. The story starts with the following remarkable theorem, due to Aissen, Endrei, Schoenberg and Whitney [2, 3].

THEOREM 4.1 ([2, 3]). *Let $(a_0,\dots,a_n)$ be a sequence of nonnegative real numbers with $a_n > 0$. The associated generating polynomial is $f(z) := \sum_{i=0}^n a_i z^i$. Then the following conditions are equivalent*:

1. *The polynomial $f(z)$ has only real roots.*
2. *The sequence $(a_0/f(1),\dots,a_n/f(1))$ is the probability distribution of a $\mathrm{PB}(p_1,\dots,p_n)$ distribution for some $p_i > 0$. The real roots of $f(z)$ are $-(1-p_i)/p_i$.*
3. *The sequence $(a_0,\dots,a_n)$ is a Pólya frequency (PF) sequence. That is, the Toeplitz matrix $(a_{j-i} : 0 \le i, j \le n)$ (with convention $a_k = 0$ for $k < 0$) is totally nonnegative: every minor of $(a_{j-i} : 0 \le i, j \le n)$ has nonnegative determinant.*

See [6] for background on total positivity. For an $n \times n$ matrix there are $\sum_{i=1}^n \binom{n}{i}^2 = \binom{2n}{n} - 1$ minors, but for the Toeplitz matrix $(a_{j-i} : 0 \le i, j \le n)$ many minors are zero. From a computational aspect, the condition (3) boils down to solving a system of $n(n-1)/2$ polynomial inequalities by matrix elimination [42, 46]. Theorem 4.1 justifies the alternative name "PF distribution" for the Poisson binomial distribution. Standard references for

PF sequences are [24, 100]. See also [84] for probabilistic interpretations for polynomials with only negative real roots, and [57] for various extensions of Theorem 4.1 by linear algebra.

EXAMPLE 3. Hypergeometric distribution. As an application of Theorem 4.1, we present a lesser known fact that the hypergeometric distribution is Poisson binomial. This result is due to Vatutin and Mikhailov [106]. The hypergeometric distribution HyperGeo(n, K, N) describes the probability of a number of successes in n draws without replacement from a population of size N that contains K successes in total. It is straightforward from this definition that a hypergeometric random variable can be written as a sum of identically distributed but dependent Bernoulli's. It is less obvious that a hypergeometric random variable can also be expressed as a sum of independent but not identically distributed Bernoulli's. We follow a simple argument in [66].

For $X \sim$ HyperGeo(n, K, N), the probability mass function of X is

$$\mathbb{P}(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \quad \text{for } 0 \le k \le K. \tag{17}$$

By simple algebra, the probability mass function (17) can also be written as

$$\mathbb{P}(X = k) = \frac{\binom{n}{k}\binom{N-n}{K-k}}{\binom{N}{K}} \quad \text{for } 0 \le k \le K,$$

thus, the PGF of X is

$$f(z) = \frac{1}{\binom{N}{K}} \sum_{k=0}^{K} \binom{n}{k}\binom{N-n}{K-k} z^k. \tag{18}$$

The key idea is to relate the PGF (18) to the Jacobi polynomials defined by

$$\begin{aligned} P_K^{a,b}(x) := \frac{1}{2^K} \sum_{k=0}^{K} \binom{K+a}{k}\binom{K+b}{K-k} \\ \times (x-1)^{K-k}(x+1)^k, \end{aligned} \tag{19}$$

for $-1 < x < 1$, where $a, b > -1$. It is well known that for any fixed a, b, the family $(P_K^{a,b}, K \ge 0)$ form orthogonal polynomials, and hence each polynomial $P_K^{a,b}$ has K real roots in $(-1, 1)$. By setting $a = n - K$, $b = N - n - K$, we have

$$\frac{f(z)}{(z-1)^K}\bigg|_{z=\frac{x+1}{x-1}} = \frac{1}{\binom{N}{n}} P_K^{a,b}(x). \tag{20}$$

The identity (20) implies that the PGF f has K roots in $(-\infty, 0)$, and the conclusion then follows Theorem 4.1.

A polynomial is called stable if it has no roots with positive imaginary part, and a stable polynomial with all real coefficients is called real stable [21, 22]. In [23], a discrete distribution on a subset of nonnegative integers is said to be strong Rayleigh if its PGF is real stable. It was also shown that the strong Rayleigh property enjoys all virtues of negative dependence. The following result is a simple consequence of Theorem 4.1.

COROLLARY 4.2. *A random variable $X \sim$ PB$(p_1, \dots, p_n)$ for some p_i if and only if X is strong Rayleigh on $\{0, \dots, n\}$.*

In the sequel, we use the terminologies "Poisson binomial" and "strong Rayleigh" interchangeably. Call a polynomial $f(z) = \sum_{i=0}^{n} a_i z^i$ with $a_i \ge 0$ strong Rayleigh if it satisfies one of the conditions in Theorem 4.1.

For $n \ge 5$, it is hopeless to get any "simple" necessary and sufficient condition for a polynomial f to be strong Rayleigh due to Abel's impossibility theorem. The most obvious necessary condition for a polynomial f to be strong Rayleigh is the Newton's inequality:

$$\begin{aligned} a_i^2 \ge a_{i-1}a_{i+1}\left(1 + \frac{1}{i}\right)\left(1 + \frac{1}{n-i}\right), \\ 1 \le i \le n-2. \end{aligned} \tag{21}$$

The sequence $(a_i; 0 \le i \le n)$ satisfying (21) is also said to be ultra-logconcave [81]. Consequently, $(a_i; 0 \le i \le n)$ is logconcave and unimodal. A lesser known sufficient condition is given in [60, 68]:

$$a_i^2 > 4a_{i-1}a_{i+1}, \quad 1 \le i \le n-2. \tag{22}$$

See also [51, 64] for various generalizations. As observed in [65], the inequality (22) cannot be improved since the sequence $(m_i; i \ge 0)$ defined by

$$m_i := \inf\left\{\frac{a_i^2}{a_{i-1}a_{i+1}}; f \text{ is strong Rayleigh}\right\},$$

decreases from $m_1 = 4$ to its limit approximately 3.2336.

Recently, determinantal point processes (DPPs) have become a useful tool to model the phenomenon of negative dependence in data diversity [48, 69], sampling [4, 78] and machine learning [1, 67]. More precisely, a DPP is a simple point process on a suitably nice space Λ, whose correlation functions at $(x_1, \dots, x_n) \in \Lambda^n$ (i.e., the probability densities of having random points at $x_1, \dots, x_n$) are

$$\rho(x_1, \dots, x_n) = \det\bigl[K(x_i, x_j)\bigr]_{1 \le i,j \le n},$$

for some kernel $K : \Lambda \times \Lambda \to \mathbb{R}$. In the context of recommender systems, let Λ be a universe of items. The negative dependence is captured by the kernel K which defines a measure of similarity between pairs of items, so that similar items are less likely to co-occur. That is, DPPs assign higher probability to sets of items that are diverse: DPPs recommend items covering various aspects of user demands rather than proposing the most popular ones.

There is a growing interest in understanding statistical properties of DPPs (see, e.g., [48]), and one of the most important quantities is the occupation counts of DPPs. For $S_1^{(n)}, \dots, S_d^{(n)}$ disjoint Borel sets indexed by $n \ge 1$, let $X_{n,i}$ be the number of points of the DPP in $S_i^{(n)}$, $1 \le i \le d$. It was proved in [99] that under fairly general conditions on the kernel K and $S_1^{(n)}, \dots, S_d^{(n)}$, the vector of counts $(X_{n,1}, \dots, X_{n,d})$ has a multivariate Gaussian limit after suitable scaling. This is equivalent to the convergence in law:

$$\frac{\sum_{i=1}^d \alpha_i X_{n,i} - \mathbb{E}(\sum_{i=1}^d \alpha_i X_{n,i})}{\mathrm{Var}(\sum_{i=1}^d \alpha_i X_{n,i})} \xrightarrow{d} \mathcal{N}(0,1), \tag{23}$$

for any real numbers $\alpha_1, \dots, \alpha_d$. The proof of the convergence in law (23) relies on the fact that the kth cumulant of the left side in (23) goes to 0 for all $k > 2$. But this proof is quite specific to the determinantal structure, and is hard to generalize to other point processes.

Let

$$\begin{aligned} &f(z_1, \dots, z_d) \\ &:= \mathbb{E}\prod_{i=1}^d z_i^{X_i} \\ &= \sum_{k_i \in \{0,1,\dots,\}} \mathbb{P}(X_i = k_i, 1 \le i \le d) \prod_{i=1}^n z_i^{k_i}, \end{aligned} \tag{24}$$

be the multivariate PGF of the counts $(X_1, \dots, X_d)$ of a DPP. It is well known that if each X_i is bounded, f is real stable in the sense that $f = 0$ has no solution $(z_1, \dots, z_d)$ with each z_i having positive imaginary part. Similar to the univariate case, a nonnegative integer-valued random vector $(X_1, \dots, X_d)$ with real stable PGF is said to be strong Rayleigh. Thus, the occupation counts of DPPs are strong Rayleigh. A natural idea is to extend the multivariate CLT of the occupation counts of DPPs to those of more general point processes which satisfy the strong Rayleigh property. This program was carried out in [47], which we call the Ghosh–Liggett–Pemantle (GLP) principle.

GLP PRINCIPLE. Strong Rayleigh vectors have multivariate Gaussian limits.

To be more precise, for $(X_{n,1}, \dots, X_{n,d}) \in \{0, \dots, n\}^d$ a sequence of strong Rayleigh vectors, let Σ_n be its covariance matrix, and σ_n^2 be the ℓ_2-norm of Σ_n. It was proved in [47] that under the conditions $\sigma_n \to \infty$, $\Sigma_n/\sigma_n^2 \to \Sigma^*$ and $\sigma_n/n^{\frac{1}{3}} \to \infty$,

$$\begin{aligned} &\frac{(X_{n,1}, \dots, X_{n,d}) - \mathbb{E}(X_{n,1}, \dots, X_{n,d})}{\sigma_n} \\ &\quad \xrightarrow{d} \mathcal{N}(0, \Sigma^*), \end{aligned} \tag{25}$$

where $\mathcal{N}(0, \Sigma^*)$ denotes the multivariate Gaussian distribution with mean 0 and covariance matrix Σ^*. The convergence in law (25) is a multivariate CLT, and its univariate counterpart was used to prove the asymptotic normality in some combinatorial problems [52, 106]. By a version of the Crámer-Wold device (see [47], Corollary 6.3'), it boils down to proving that for any positive rational numbers $q_1, \dots, q_d$, the linear combination $\sum_{i=1}^d q_i X_{n,i}$ has a Gaussian limit law. This follows from the fact that the PGF of the random variable $\sum_{i=1}^d q_i X_{n,i}$ has no zeros near 1, which is a consequence of the strong Rayleigh property of $(X_{n,1}, \dots, X_{n,d})$. The condition $\sigma_n/n^{\frac{1}{3}} \to \infty$ (i.e., the variance grows fast) is purely technical, and it has been removed later in [73, 74]. However, these works rely heavily on analytical tools in complex analysis.

Here we outline a probabilistic argument to prove the multivariate CLT (25), which motivates the study of rational multiples of strong Rayleigh, or Poisson binomial variables. In light of [52, 106], the proof is complete if one can approximate $\sum_{i=1}^d q_i X_{n,i}$ by a strong Rayleigh variable. We consider the univariate subcase by taking $q_i = 0$, $i \ge 2$: if X is strong Rayleigh, how well can one approximate jX/k for each $j, k \ge 1$ by a strong Rayleigh variable? A good approximation may serve as a starting point to a probabilistic proof of the multivariate CLT (25) for strong Rayleigh distributions. The case $j = 1$ was solved in [47].

THEOREM 4.3 ([47]). *Let X be strong Rayleigh. Then $\lfloor \frac{X}{k} \rfloor$ is strong Rayleigh for each $k \ge 1$.*

The key to the proof of Theorem 4.3 is that for f a polynomial of degree n and $k \ge 1$, write $f(z) = \sum_{j=0}^{k-1} x^j g_j(z^k)$, with g_j a polynomial of degree $\lfloor \frac{n-j}{k} \rfloor$. The theorem asserts that if f is strong Rayleigh, then each g_i is real-rooted, and their roots are interlaced in the sense that if the set of all $n-k+1$ roots z_j of the g_i's are placed in increasing order $z_{n-k} < \cdots < z_1 < z_0 < 0$, then the roots of g_i are $z_i, z_{i+k}, z_{i+2k} \dots$ In fact, the real-rootedness follows from the fact that $(a_n; n \ge 0)$ is a PF sequence implies $(a_{kn+j}; n \ge 0)$ is a PF sequence for each $k \ge 1$ and $0 \le j < k$. This result is well known, see [2, 24]. But the root interlacing seems less obvious by PF sequences.

A natural question is whether $\lfloor jX/k \rfloor$ is strong Rayleigh for each $j, k \ge 1$. It turns out that $\lfloor 2X/3 \rfloor$ can be far away from being strong Rayleigh. In fact, one can prove that for $X \sim \mathrm{Bin}(3n, 1/2)$, and z_i being the roots of the PGF of $\lfloor 2X/3 \rfloor$, $\max_i\{\Im(z_i)\} \ge \sqrt{\frac{9n^2-9n-1}{2}}$ where $\Im(z)$ is the imaginary part of z. The reason why some roots of the PGF of $\lfloor 2X/3 \rfloor$ have large positive imaginary parts is due to the unbalanced allocation of probability weights to even and odd numbers: $\mathbb{P}(\lfloor \frac{2X}{3} \rfloor = 2k) = \binom{3n+1}{3k+1}$ while $\mathbb{P}(\lfloor \frac{2X}{3} \rfloor = 2k+1) = \binom{3n}{3k+2}$. So the Newton's inequality (21) is not satisfied. In the Supplementary Materials [102], we formulate the problem of approximating rational fractions of Poisson binomial via optimal transport, and provide some analysis of $\frac{2X}{3}$ for X a binomial random variable.

Recently, Liggett proved an interesting result of $\lfloor 2X/3 \rfloor$ for X a strong Rayleigh variable.

THEOREM 4.4 ([71]). *Let X be strong Rayleigh. Then the PGF of* $\lfloor 2X/3 \rfloor$ *is Hurwitz stable. That is, all its roots have negative real parts.*

The idea is to write the PGF of $\lfloor 2X/3 \rfloor$ as $g_0(x^2) + xg_1(x^2)$, where g_0 and g_1 have interlacing roots. By the Hermite–Biehler theorem [12, 53], such polynomials are Hurwitz stable. This means that the PGF of $\lfloor 2X/3 \rfloor$ can be factorized into polynomials with positive coefficients of degrees no greater than 2. Thus, $\lfloor 2X/3 \rfloor$ is a Poisson multinomial variable, that is the sum of independent random variables with values in $\{0, 1, 2\}$. The following result is conjectured.

CONJECTURE 4.5. *Let X be strong Rayleigh. Then* $\lfloor jX/k \rfloor$ *is the sum of independent random variables with values in* $\{0, 1, \dots, j\}$. *Equivalently, the PGF of* $\lfloor jX/k \rfloor$ *can be factorized into polynomials with positive coefficients of degrees no greater than* j.

5. COMPUTATIONS OF POISSON BINOMIAL DISTRIBUTIONS

In this section, we discuss a few computational issues of learning and computing the Poisson binomial distribution.

Learning the Poisson binomial distribution. Distribution learning is an active domain in both statistics and computer science. Following [36], given access to independent samples from an unknown distribution P, an error control $\epsilon > 0$ and a confidence level $\delta > 0$, a learning algorithm outputs an estimation $\widehat{P}$ such that $\mathbb{P}(d_{\mathrm{TV}}(\widehat{P}, P) \le \epsilon) \ge 1 - \delta$. The performance of a learning algorithm is measured by its sample complexity and its computational complexity.

For $X \sim \mathrm{PB}(p_1, \dots, p_n)$, this amounts to finding a vector $(\widehat{p}_1, \dots, \widehat{p}_n)$ defining $\widehat{X} \sim \mathrm{PB}(\widehat{p}_1, \dots, \widehat{p}_n)$ such that $d_{\mathrm{TV}}(\widehat{X}, X)$ is small with high probability. This is often called proper learning of Poisson binomial distributions. Building upon previous work [14, 35, 89], Daskalakis, Diakonikolas and Servedio [34] established the following result for proper learning of Poisson binomial distributions.

THEOREM 5.1 ([34]). *Let* $X \sim \mathrm{PB}(p_1, \dots, p_n)$ *with unknown* p_i*'s. There is an algorithm such that given* $\epsilon, \delta > 0$, *it requires*:

- (*sample complexity*) $O(1/\epsilon^2) \cdot \log(1/\delta)$ *independent samples from* X,
- (*computational complexity*) $(1/\epsilon)^{O(\log^2(1/\epsilon))} \cdot O(\log n \cdot \log(1/\delta))$ *operations*,

to construct a vector $(\widehat{p}_1, \dots, \widehat{p}_n)$ *satisfying* $\mathbb{P}(d_{\mathrm{TV}}(\widehat{X}, X) \le \epsilon) \ge 1 - \delta$ *for* $\widehat{X} \sim \mathrm{PB}(\widehat{p}_1, \dots, \widehat{p}_n)$.

The proof of Theorem 5.1 relies on the fact each Poisson binomial distribution is either close to a Poisson binomial distribution whose support is sparse, or is close to a translated "heavy" binomial distribution. The key to the algorithm is to find subsets covering all Poisson binomial distributions, and each of these subsets is either "sparse" or "heavy." Applying Birgé's unimodal algorithm [14] to sparse subsets, and the translated Poisson approximation (Theorem 3.2) to heavy subsets give the desired algorithm. Note that the sample complexity in Theorem 5.1 is nearly optimal, since $\Theta(1/\epsilon^2)$ samples are required to distinguish $\mathrm{Bin}(n, 1/2)$ from $\mathrm{Bin}(n, 1/2 + \epsilon/\sqrt{n})$ which differ by $\Theta(\epsilon)$ in total variation. See also [37] for further results on learning the Poisson binomial distribution, and [33, 38, 39] for the integer-valued distribution.

Computing the Poisson binomial distribution. Recall the probability distribution of $X \sim \mathrm{PB}(p_1, \dots, p_n)$ from (1). Given $p_1, \dots, p_n$, a brute-force computation of this distribution is expensive for large n. Approximations in Section 3 are often used to estimate the probability distribution/CDF of the Poisson binomial distribution. Here we focus on the efficient algorithms to compute exactly these distribution functions. There are two general approaches: recursive formulas and discrete Fourier analysis.

In [29], the authors presented several recursive algorithms to compute (1). For $0 \le k \le m \le n$, let

$$R_{k,m} := \mathbb{P}(X_m = k), \quad \text{where } X_m \sim \mathrm{PB}(p_1, \dots, p_m).$$

So $\mathbb{P}(X = k) = R_{k,n}$ for $0 \le k \le n$. Two recursive algorithms are proposed:

- [45] For $0 \le k \le m \le n$,

$$R_{k,m} = (1 - p_m)R_{k,m-1} + p_m R_{k-1,m-1}, \tag{26}$$

  with the convention that $R_{-1,m} = R_{m+1,m} = 0$ for $0 \le m \le n-1$, and $R_{0,0} = 1$.
- [30, 101] For $i \ge 1$, let $t_i := \sum_{j=1}^n (\frac{p_j}{1-p_j})^i$. Then

$$\begin{aligned} R_{0,n} &= \prod_{j=1}^n (1 - p_j), \\ R_{k,n} &= \frac{1}{k} \sum_{i=1}^k (-1)^{i+1} t_i R_{k-i,n}, \quad 1 \le k \le n. \end{aligned} \tag{27}$$

The recursion (26) uses 1 addition and 2 multiplications to compute $R_{k,m}$ given $R_{k,m-1}$ and $R_{k-1,m-1}$. Thus, to get the value of $R_{k,n}$ for a general k, it requires $3k(n-k)$ operations. For the recursion (27), if the values of t_i are available, it uses $k-1$ additions and $k+1$ multiplications to compute $R_{k,n}$ given $R_{0,n}, \dots, R_{k-1,n}$. Moreover, it takes $n-1$ additions to get t_1, and $n-1$ additions and n multiplications to get t_i for $i \ge 2$. As a result, it requires approximately $k^2 + 2kn$ operations to get the value of $R_{k,n}$ for any k. So for small k's, the recursion (27) uses less operations than the recursion (26) to compute the value of $R_{k,n}$. However, to output the whole sequence of $\{R_{0,n}, \dots, R_{n,n}\}$ (with memory being allowed), the recursion (26) requires approximately $3n^2/2$ operations, and

n in memory while the recursion (27) requires approximately $3n^2$ operations, and n in memory.

In another direction, [43, 58] used a Fourier approach to evaluate the probability distribution/CDF of Poisson binomial distributions. They provided the following explicit formulas:

$$\mathbb{P}(X = k) = \frac{1}{n+1} \sum_{j=0}^{n} \exp(-i\omega kj) x_j, \tag{28}$$

and

$$\mathbb{P}(X \le k) = \frac{1}{n+1} \sum_{j=0}^{n} \frac{1 - \exp(-i\omega(k+1)j)}{1 - \exp(-i\omega j)} x_j, \tag{29}$$

where $\omega := \frac{2\pi}{n+1}$ and $x_j := \prod_{k=1}^{n}(1 - p_k + p_k \exp(i\omega j))$. Note that it uses n operations to get x_0, and $6(n-1)$ operations to get x_j for $j \neq 0$. Since x_j and x_{n+1-j} are conjugate of each other, it requires approximately $3n^2$ operations to compute the sequence $\{x_0, \dots, x_n\}$. Further, the r.h.s. of (28) is the discrete Fourier transform of $\{x_0, \dots, x_n\}$ which can be easily computed by Fast Fourier Transform in $\mathcal{O}(n \log n)$ operations. See also [15] for a related approach.

ACKNOWLEDGMENTS

We thank Tom Liggett, Jim Pitman and Terry Tao for helpful discussions. We thank Yuting Ye for providing Example 2, and C.-H. Zhang for the simple proof in Example 1(3). We also thank an anonymous Associate Editor and the Editors for suggestions that greatly improve the presentation of the paper.

FUNDING

The first author was supported by NSF Grant DMS-2113779 and a start-up grant at Columbia University.

SUPPLEMENTARY MATERIAL

Rational Fractions of Poisson Binomial (DOI: 10.1214/22-STS852SUPP; .pdf). In the Supplementary Material, we formulate the problem of approximating rational fractions of Poisson binomial via optimal transport, and record some analysis of $\frac{2X}{3}$ for X a binomial random variable.

REFERENCES

[1] AFFANDI, R. H., FOX, E., ADAMS, R. and TASKAR, B. (2014). Learning the parameters of determinantal point process kernels. In *International Conference on Machine Learning* 1224–1232.

[2] AISSEN, M., EDREI, A., SCHOENBERG, I. J. and WHITNEY, A. (1951). On the generating functions of totally positive sequences. *Proc. Natl. Acad. Sci. USA* **37** 303–307. MR0041897 https://doi.org/10.1073/pnas.37.5.303

[3] AISSEN, M., SCHOENBERG, I. J. and WHITNEY, A. M. (1952). On the generating functions of totally positive sequences. I. *J. Anal. Math.* **2** 93–103. MR0053174 https://doi.org/10.1007/BF02786970

[4] ANARI, N., GHARAN, S. O. and REZAEI, A. (2016). Monte Carlo Markov chain algorithms for sampling strongly Rayleigh distributions and determinantal point processes. In *Conference on Learning Theory* 103–115.

[5] ANDERSON, T. W. and SAMUELS, S. M. (1967). Some inequalities among binomial and Poisson probabilities. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability* (*Berkeley, Calif.*, 1965/66) 1–12. Univ. California Press, Berkeley, CA. MR0214176

[6] ANDO, T. (1987). Totally positive matrices. *Linear Algebra Appl.* **90** 165–219. MR0884118 https://doi.org/10.1016/0024-3795(87)90313-2

[7] AZUMA, K. (1967). Weighted sums of certain dependent random variables. *Tohoku Math. J.* (2) **19** 357–367. MR0221571 https://doi.org/10.2748/tmj/1178243286

[8] BARBOUR, A. D. and HALL, P. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.* **95** 473–480. MR0755837 https://doi.org/10.1017/S0305004100061806

[9] BARBOUR, A. D., HOLST, L. and JANSON, S. (1992). *Poisson Approximation. Oxford Studies in Probability* **2**. The Clarendon Press, Oxford University Press, New York. MR1163825

[10] BEREND, D. and TASSA, T. (2010). Improved bounds on Bell numbers and on moments of sums of random variables. *Probab. Math. Statist.* **30** 185–205. MR2792580

[11] BICKEL, P. J. and VAN ZWET, W. R. (1980). On a theorem of Hoeffding. In *Asymptotic Theory of Statistical Tests and Estimation* (*Proc. Adv. Internat. Sympos., Univ. North Carolina, Chapel Hill, N.C.*, 1979) 307–324. Academic Press, New York-Toronto, Ont. MR0571346

[12] BIEHLER, R. (1879). Sur une classe d'équations algébriques dont toutes les racines sont réelles. *J. Reine Angew. Math.* **87** 350–352. MR1579799 https://doi.org/10.1515/crll.1879.87.350

[13] BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR1324786

[14] BIRGÉ, L. (1997). Estimation of unimodal densities without smoothness assumptions. *Ann. Statist.* **25** 970–981. MR1447736 https://doi.org/10.1214/aos/1069362733

[15] BISCARRI, W., ZHAO, S. D. and BRUNNER, R. J. (2018). A simple and fast method for computing the Poisson binomial distribution function. *Comput. Statist. Data Anal.* **122** 92–100. MR3765817 https://doi.org/10.1016/j.csda.2018.01.007

[16] BOBKOV, S. G. (2018). Berry–Esseen bounds and Edgeworth expansions in the central limit theorem for transport distances. *Probab. Theory Related Fields* **170** 229–262. MR3748324 https://doi.org/10.1007/s00440-017-0756-2

[17] BOLAND, P. J. (2007). The probability distribution for the number of successes in independent trials. *Comm. Statist. Theory Methods* **36** 1327–1331. MR2396538 https://doi.org/10.1080/03610920601077014

[18] BOLAND, P. J. and PROSCHAN, F. (1983). The reliability of k out of n systems. *Ann. Probab.* **11** 760–764. MR0704562

[19] BOLAND, P. J., SINGH, H. and CUKIC, B. (2002). Stochastic orders in partition and random testing of software. *J. Appl. Probab.* **39** 555–565. MR1928890 https://doi.org/10.1239/jap/1034082127

[20] BOLAND, P. J., SINGH, H. and CUKIC, B. (2004). The stochastic precedence ordering with applications in sampling and testing. *J. Appl. Probab.* **41** 73–82. MR2036272 https://doi.org/10.1239/jap/1077134668

[21] BORCEA, J. and BRÄNDÉN, P. (2008). Applications of stable polynomials to mixed determinants: Johnson's conjectures, unimodality, and symmetrized Fischer products. *Duke Math. J.* **143** 205–223. MR2420507 https://doi.org/10.1215/00127094-2008-018

[22] BORCEA, J. and BRÄNDÉN, P. (2009). Pólya–Schur master theorems for circular domains and their boundaries. *Ann. of Math.* (2) **170** 465–492. MR2521123 https://doi.org/10.4007/annals.2009.170.465

[23] BORCEA, J., BRÄNDÉN, P. and LIGGETT, T. M. (2009). Negative dependence and the geometry of polynomials. *J. Amer. Math. Soc.* **22** 521–567. MR2476782 https://doi.org/10.1090/S0894-0347-08-00618-8

[24] BRENTI, F. (1989). Unimodal, log-concave and Pólya frequency sequences in combinatorics. *Mem. Amer. Math. Soc.* **81** viii+106. MR0963833 https://doi.org/10.1090/memo/0413

[25] BRODERICK, T., PITMAN, J. and JORDAN, M. I. (2013). Feature allocations, probability functions, and paintboxes. *Bayesian Anal.* **8** 801–836. MR3150470 https://doi.org/10.1214/13-BA823

[26] CHATTERJEE, S., DIACONIS, P. and MECKES, E. (2005). Exchangeable pairs and Poisson approximation. *Probab. Surv.* **2** 64–106. MR2121796 https://doi.org/10.1214/154957805100000096

[27] CHEN, L. H. Y. and SHAO, Q.-M. (2001). A non-uniform Berry–Esseen bound via Stein's method. *Probab. Theory Related Fields* **120** 236–254. MR1841329 https://doi.org/10.1007/PL00008782

[28] CHEN, L. H. Y. and SHAO, Q.-M. (2005). Stein's method for normal approximation. In *An Introduction to Stein's Method. Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **4** 1–59. Singapore Univ. Press, Singapore. MR2235448 https://doi.org/10.1142/9789812567680_0001

[29] CHEN, S. X. and LIU, J. S. (1997). Statistical applications of the Poisson-binomial and conditional Bernoulli distributions. *Statist. Sinica* **7** 875–892. MR1488647

[30] CHEN, X.-H., DEMPSTER, A. P. and LIU, J. S. (1994). Weighted finite population sampling to maximize entropy. *Biometrika* **81** 457–469. MR1311090 https://doi.org/10.1093/biomet/81.3.457

[31] CHOI, K. P. and XIA, A. (2002). Approximating the number of successes in independent trials: Binomial versus Poisson. *Ann. Appl. Probab.* **12** 1139–1148. MR1936586 https://doi.org/10.1214/aoap/1037125856

[32] DARROCH, J. N. (1964). On the distribution of the number of successes in independent trials. *Ann. Math. Stat.* **35** 1317–1321. MR0164359 https://doi.org/10.1214/aoms/1177703287

[33] DASKALAKIS, C., DIAKONIKOLAS, I., O'DONNELL, R., SERVEDIO, R. A. and TAN, L.-Y. (2013). Learning sums of independent integer random variables. In 2013 *IEEE* 54*th Annual Symposium on Foundations of Computer Science—FOCS* 2013 217–226. IEEE Computer Soc., Los Alamitos, CA. MR3246223 https://doi.org/10.1109/FOCS.2013.31

[34] DASKALAKIS, C., DIAKONIKOLAS, I. and SERVEDIO, R. A. (2015). Learning Poisson binomial distributions. *Algorithmica* **72** 316–357. MR3332935 https://doi.org/10.1007/s00453-015-9971-3

[35] DASKALAKIS, C. and PAPADIMITRIOU, C. (2015). Sparse covers for sums of indicators. *Probab. Theory Related Fields* **162** 679–705. MR3383340 https://doi.org/10.1007/s00440-014-0582-8

[36] DEVROYE, L. and LUGOSI, G. (2001). *Combinatorial Methods in Density Estimation. Springer Series in Statistics*. Springer, New York. MR1843146 https://doi.org/10.1007/978-1-4613-0125-7

[37] DIAKONIKOLAS, I., KANE, D. M. and STEWART, A. (2016). Properly learning Poisson binomial distributions in almost polynomial time. In *Conference on Learning Theory* 850–878.

[38] DIAKONIKOLAS, I., KANE, D. M. and STEWART, A. (2016). The Fourier transform of Poisson multinomial distributions and its algorithmic applications. In *STOC'*16*—Proceedings of the* 48*th Annual ACM SIGACT Symposium on Theory of Computing* 1060–1073. ACM, New York. MR3536636 https://doi.org/10.1145/2897518.2897552

[39] DIAKONIKOLAS, I., KANE, D. M. and STEWART, A. (2016). Optimal learning via the Fourier transform for sums of independent integer random variables. In *Conference on Learning Theory* 831–849.

[40] DUFFIE, D., SAITA, L. and WANG, K. (2007). Multi-period corporate default prediction with stochastic covariates. *J. Financ. Econ.* **83** 635–665.

[41] EHM, W. (1991). Binomial approximation to the Poisson binomial distribution. *Statist. Probab. Lett.* **11** 7–16. MR1093412 https://doi.org/10.1016/0167-7152(91)90170-V

[42] FALLAT, S., JOHNSON, C. R. and SOKAL, A. D. (2017). Total positivity of sums, Hadamard products and Hadamard powers: Results and counterexamples. *Linear Algebra Appl.* **520** 242–259. MR3611466 https://doi.org/10.1016/j.laa.2017.01.013

[43] FERNÁNDEZ, M. and WILLIAMS, S. (2010). Closed-form expression for the Poisson-binomial probability density function. *IEEE Trans. Aerosp. Electron. Syst.* **46** 803–817.

[44] FERNANDEZ, M. F. and ARIDGIDES, T. (2003). Measures for evaluating sea mine identification processing performance and the enhancements provided by fusing multisensor/multiprocess data via an M-out-of-N voting scheme. In *Detection and Remediation Technologies for Mines and Minelike Targets VIII* **5089** 425–436.

[45] GAIL, M. H., LUBIN, J. H. and RUBINSTEIN, L. V. (1981). Likelihood calculations for matched case-control studies and survival studies with tied death times. *Biometrika* **68** 703–707. MR0637792 https://doi.org/10.1093/biomet/68.3.703

[46] GASCA, M. and PEÑA, J. M. (1992). Total positivity and Neville elimination. *Linear Algebra Appl.* **165** 25–44. MR1149743 https://doi.org/10.1016/0024-3795(92)90226-Z

[47] GHOSH, S., LIGGETT, T. M. and PEMANTLE, R. (2017). Multivariate CLT follows from strong Rayleigh property. In 2017 *Proceedings of the Fourteenth Workshop on Analytic Algorithmics and Combinatorics* (*ANALCO*) 139–147. SIAM, Philadelphia, PA. MR3630942 https://doi.org/10.1137/1.9781611974775.14

[48] GHOSH, S. and RIGOLLET, P. (2020). Gaussian determinantal processes: A new model for directionality in data. *Proc. Natl. Acad. Sci. USA* **117** 13207–13213. MR4240249 https://doi.org/10.1073/pnas.1917151117

[49] GLESER, L. J. (1975). On the distribution of the number of successes in independent trials. *Ann. Probab.* **3** 182–188. MR0365651 https://doi.org/10.1214/aop/1176996461

[50] GOLDSTEIN, L. (2010). Bounds on the constant in the mean central limit theorem. *Ann. Probab.* **38** 1672–1689. MR2663641 https://doi.org/10.1214/10-AOP527

[51] HANDELMAN, D. (2013). Arguments of zeros of highly log concave polynomials. *Rocky Mountain J. Math.* **43** 149–177. MR3065459 https://doi.org/10.1216/RMJ-2013-43-1-149

[52] HARPER, L. H. (1967). Stirling behavior is asymptotically normal. *Ann. Math. Stat.* **38** 410–414. MR0211432 https://doi.org/10.1214/aoms/1177698956

[53] HERMITE, C. (1856). Extrait d'une lettre de Mr. Ch. Hermite de Paris à Mr. Borchardt de Berlin sur le nombre des

racines d'une équation algébrique comprises entre des limites données. *J. Reine Angew. Math.* **52** 39–51. MR1578969 https://doi.org/10.1515/crll.1856.52.39
[54] HILLION, E. and JOHNSON, O. (2017). A proof of the Shepp–Olkin entropy concavity conjecture. *Bernoulli* **23** 3638–3649. MR3654818 https://doi.org/10.3150/16-BEJ860
[55] HOEFFDING, W. (1956). On the distribution of the number of successes in independent trials. *Ann. Math. Stat.* **27** 713–721. MR0080391 https://doi.org/10.1214/aoms/1177728178
[56] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363
[57] HOLTZ, O. and TYAGLOV, M. (2012). Structured matrices, continued fractions, and root localization of polynomials. *SIAM Rev.* **54** 421–509. MR2966723 https://doi.org/10.1137/090781127
[58] HONG, Y. (2013). On computing the distribution function for the Poisson binomial distribution. *Comput. Statist. Data Anal.* **59** 41–51. MR3000040 https://doi.org/10.1016/j.csda.2012.10.006
[59] HONG, Y., MEEKER, W. Q. and MCCALLEY, J. D. (2009). Prediction of remaining life of power transformers based on left truncated and right censored lifetime data. *Ann. Appl. Stat.* **3** 857–879. MR2750685 https://doi.org/10.1214/00-AOAS231
[60] HUTCHINSON, J. I. (1923). On a remarkable class of entire functions. *Trans. Amer. Math. Soc.* **25** 325–332. MR1501248 https://doi.org/10.2307/1989293
[61] JANSON, S. (1994). Coupling and Poisson approximation. *Acta Appl. Math.* **34** 7–15. MR1273843 https://doi.org/10.1007/BF00994254
[62] JOGDEO, K. and SAMUELS, S. M. (1968). Monotone convergence of binomial probabilities and a generalization of Ramanujan's equation. *Ann. Math. Stat.* **39** 1191–1195.
[63] KARR, A. F. (1993). *Probability. Springer Texts in Statistics*. Springer, New York. MR1231974 https://doi.org/10.1007/978-1-4612-0891-4
[64] KATKOVA, O. M. and VISHNYAKOVA, A. M. (2008). A sufficient condition for a polynomial to be stable. *J. Math. Anal. Appl.* **347** 81–89. MR2433826 https://doi.org/10.1016/j.jmaa.2008.05.079
[65] KOSTOV, V. P. and SHAPIRO, B. (2013). Hardy–Petrovitch–Hutchinson's problem and partial theta function. *Duke Math. J.* **162** 825–861. MR3047467 https://doi.org/10.1215/00127094-2087264
[66] KOU, S. G. and YING, Z. (1996). Asymptotics for a 2×2 table with fixed margins. *Statist. Sinica* **6** 809–829. MR1422405
[67] KULESZA, A. and TASKAR, B. (2012). Determinantal point processes for machine learning. *Found. Trends Mach. Learn.* **5** 123–286.
[68] KURTZ, D. C. (1992). A sufficient condition for all the roots of a polynomial to be real. *Amer. Math. Monthly* **99** 259–263. MR1216215 https://doi.org/10.2307/2325063
[69] LAVANCIER, F., MØLLER, J. and RUBAK, E. (2015). Determinantal point process models and statistical inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 853–877. MR3382600 https://doi.org/10.1111/rssb.12096
[70] LE CAM, L. (1960). An approximation theorem for the Poisson binomial distribution. *Pacific J. Math.* **10** 1181–1197. MR0142174
[71] LIGGETT, T. M. (2021). Approximating multiples of strong Rayleigh random variables. Available at https://celebratio.org/Liggett_T/article/858/.
[72] MARSHALL, A. W., OLKIN, I. and ARNOLD, B. C. (2011). *Inequalities: Theory of Majorization and Its Applications*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2759813 https://doi.org/10.1007/978-0-387-68276-1
[73] MICHELEN, M. and SAHASRABUDHE, J. (2019). Central limit theorems and the geometry of polynomials. Available at arXiv:1908.09020.
[74] MICHELEN, M. and SAHASRABUDHE, J. (2019). Central limit theorems from the roots of probability generating functions. *Adv. Math.* **358** 106840. MR4021875 https://doi.org/10.1016/j.aim.2019.106840
[75] NEAMMANEE, K. (2005). On the constant in the nonuniform version of the Berry–Esseen theorem. *Int. J. Math. Math. Sci.* **12** 1951–1967. MR2176447 https://doi.org/10.1155/IJMMS.2005.1951
[76] NEDELMAN, J. and WALLENIUS, T. (1986). Bernoulli trials, Poisson trials, surprising variances, and Jensen's inequality. *Amer. Statist.* **40** 286–289. MR0866910 https://doi.org/10.2307/2684605
[77] NOVAK, S. Y. (2019). Poisson approximation. *Probab. Surv.* **16** 228–276. MR3992498 https://doi.org/10.1214/18-PS318
[78] OVEIS GHARAN, S., SABERI, A. and SINGH, M. (2011). A randomized rounding approach to the traveling salesman problem. In 2011 *IEEE* 52*nd Annual Symposium on Foundations of Computer Science—FOCS* 2011 550–559. IEEE Computer Soc., Los Alamitos, CA. MR2932731 https://doi.org/10.1109/FOCS.2011.80
[79] PADITZ, L. (1989). On the analytical structure of the constant in the nonuniform version of the Esseen inequality. *Statistics* **20** 453–464. MR1012316 https://doi.org/10.1080/02331888908802196
[80] PEKÖZ, E. A., RÖLLIN, A., ČEKANAVIČIUS, V. and SHWARTZ, M. (2009). A three-parameter binomial approximation. *J. Appl. Probab.* **46** 1073–1085. MR2582707 https://doi.org/10.1239/jap/1261670689
[81] PEMANTLE, R. (2000). Towards a theory of negative dependence. *J. Math. Phys.* **41** 1371–1390. MR1757964 https://doi.org/10.1063/1.533200
[82] PETROV, V. V. (1965). A bound for the deviation of the distribution of a sum of independent random variables from the normal law. *Dokl. Akad. Nauk SSSR* **160** 1013–1015. MR0178497
[83] PETROV, V. V. (1975). *Sums of Independent Random Variables. Ergebnisse der Mathematik und Ihrer Grenzgebiete*, *Band* 82. Springer, New York. Translated from the Russian by A. A. Brown. MR0388499
[84] PITMAN, J. (1997). Probabilistic bounds on the coefficients of polynomials with only real zeros. *J. Combin. Theory Ser. A* **77** 279–303. MR1429082 https://doi.org/10.1006/jcta.1997.2747
[85] PLATONOV, M. L. (1979). *Combinatorial Numbers of a Class of Mappings and Their Applications*. "Nauka", Moscow. MR0555244
[86] PLEDGER, G. and PROSCHAN, F. (1971). Comparisons of order statistics and of spacings from heterogeneous distributions. In *Optimizing Methods in Statistics* (*Proc. Sympos.*, *Ohio State Univ.*, *Columbus*, *Ohio*, 1971) 89–113. MR0341738
[87] POISSON, S. D. (1837). *Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile*. Bachelier.
[88] RIO, E. (2009). Upper bounds for minimal distances in the central limit theorem. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** 802–817. MR2548505 https://doi.org/10.1214/08-AIHP187
[89] RÖLLIN, A. (2007). Translated Poisson approximation using exchangeable pair couplings. *Ann. Appl. Probab.* **17** 1596–1614. MR2358635 https://doi.org/10.1214/105051607000000258
[90] ROOS, B. (1999). Asymptotic and sharp bounds in the Poisson approximation to the Poisson-binomial distribution. *Bernoulli* **5** 1021–1034. MR1735783 https://doi.org/10.2307/3318558

[91] ROOS, B. (2000). Binomial approximation to the Poisson binomial distribution: The Krawtchouk expansion. *Teor. Veroyatn. Primen.* **45** 328–344. MR1967760 https://doi.org/10.1137/S0040585X9797821X

[92] ROSENBAUM, P. R. (2002). *Observational Studies*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1899138 https://doi.org/10.1007/978-1-4757-3692-2

[93] ROSENMAN, E. (2019). Some new results for Poisson binomial models. Available at arXiv:1907.09053.

[94] SAMUELS, S. M. (1965). On the number of successes in independent trials. *Ann. Math. Stat.* **36** 1272–1278. MR0179825 https://doi.org/10.1214/aoms/1177699998

[95] SCHUMACHER, N. (1999). Binomial option pricing with non-identically distributed returns and its implications. *Math. Comput. Modelling* **29** 121–143. MR1704770 https://doi.org/10.1016/S0895-7177(99)00097-7

[96] SHEPP, L. A. and OLKIN, I. (1981). Entropy of the sum of independent Bernoulli random variables and of the multinomial distribution. In *Contributions to Probability* 201–206. Academic Press, New York. MR0618689

[97] SHIGANOV, I. S. (1982). Refinement of the upper bound of a constant in the remainder term of the central limit theorem. In *Stability Problems for Stochastic Models* (*Moscow*, 1982) 109–115. Vsesoyuz. Nauchno-Issled. Inst. Sistem. Issled., Moscow. MR0734011

[98] SKIPPER, M. (2012). A Pólya approximation to the Poisson-binomial law. *J. Appl. Probab.* **49** 745–757. MR3012097 https://doi.org/10.1239/jap/1346955331

[99] SOSHNIKOV, A. (2002). Gaussian limit for determinantal random point fields. *Ann. Probab.* **30** 171–187. MR1894104 https://doi.org/10.1214/aop/1020107764

[100] STANLEY, R. P. (1989). Log-concave and unimodal sequences in algebra, combinatorics, and geometry. In *Graph Theory and Its Applications*: *East and West* (*Jinan*, 1986). *Ann. New York Acad. Sci.* **576** 500–535. New York Acad. Sci., New York. MR1110850 https://doi.org/10.1111/j.1749-6632.1989.tb16434.x

[101] STEIN, C. (1990). Application of Newton's identities to a generalized birthday problem and to the Poisson binomial distribution. Technical Report 354, Department of Statistics, Stanford University.

[102] TANG, W. and TANG, F. (2023). Supplement to "The Poisson binomial distribution—Old & New." https://doi.org/10.1214/22-STS852SUPP

[103] TEJADA, A. and ARNOLD, J. (2011). The role of Poisson's binomial distribution in the analysis of TEM images. *Ultramicroscopy* **111** 1553–1556.

[104] THONGTHA, P. and NEAMMANEE, K. (2007). Refinement on the constants in the non-uniform version of the Berry–Esseen theorem. *Thai J. Math.* **5** 1–13. MR2407441

[105] VAN BEEK, P. (1972). An application of Fourier methods to the problem of sharpening the Berry–Esseen inequality. *Z. Wahrsch. Verw. Gebiete* **23** 187–196. MR0329000 https://doi.org/10.1007/BF00536558

[106] VATUTIN, V. A. and MIKHAĬLOV, V. G. (1982). Limit theorems for the number of empty cells in an equiprobable scheme for the distribution of particles by groups. *Theory Probab. Appl.* **27** 734–743. MR0681461

[107] WANG, Y. H. (1993). On the number of successes in independent trials. *Statist. Sinica* **3** 295–312. MR1243388

[108] XU, M. and BALAKRISHNAN, N. (2011). On the convolution of heterogeneous Bernoulli random variables. *J. Appl. Probab.* **48** 877–884. MR2884823 https://doi.org/10.1017/s0021900200008391

[109] ZHANG, C.-H. (1999). Sub-Bernoulli functions, moment inequalities and strong laws for nonnegative and symmetrized U-statistics. *Ann. Probab.* **27** 432–453. MR1681165 https://doi.org/10.1214/aop/1022677268