The Poisson Binomial Distribution—Old & New

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 $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$, $0 \leq k \leq n$. It is well known that if $n$ is large, the 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 $X \sim \text{PB}(p_1, \ldots, p_n)$ if

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

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

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

where the sum ranges over all subset of $[n] := \{1, \ldots, n\}$ 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.
2. DISTRIBUTIONAL PROPERTIES OF POISSON BINOMIAL VARIABLES

In this section, we review a few distributional properties of the Poisson binomial distribution. For \(X \sim \text{PB}(p_1, \ldots, p_n)\), we have
\[
\mu := \mathbb{E}X = n \bar{p},
\]
and
\[
\sigma^2 := \text{Var}X = n(1 - \bar{p}) - \sum_{i=1}^{n} (p_i - \bar{p})^2.
\]

where \(\bar{p} := \frac{\sum_{i=1}^{n} p_i}{n}\). It is easily seen that by keeping \(\mathbb{E}X\) (or \(\bar{p}\)) fixed, the variance of \(X\) is increasing as \(n\) is increasing, 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 first result of this kind is due to Hoeffding \([55]\), claiming that among all Poisson binomial distributions with a given structure, the binomial distribution is the most spread out.

**Theorem 2.1** (\([55]\) Hoeffding’s inequalities). Let \(X \sim \text{PB}(p_1, \ldots, p_n)\), and \(\bar{X} \sim \text{Bin}(n, \bar{p})\).

1. There are inequalities
\[
\mathbb{P}(X \leq k) \leq \mathbb{P}(\bar{X} \leq k) \quad \text{for } 0 \leq k \leq n\bar{p} - 1
\]
and
\[
\mathbb{P}(X \leq k) \geq \mathbb{P}(\bar{X} \leq k) \quad \text{for } n\bar{p} \leq k \leq n.
\]

2. For any convex function \(g: [n] \to \mathbb{R}\) in the sense that \(g(k + 2) - 2g(k + 1) + g(k) > 0, 0 \leq k \leq n - 2\), we have
\[
\mathbb{E}g(X) \leq \mathbb{E}g(\bar{X}),
\]
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 \((p_1, \ldots, p_n) = (0, \frac{\lambda}{n - 1}, \ldots, \frac{\lambda}{n - 1})\), we get for \(X \sim \text{Bin}(n - 1, \frac{\lambda}{n - 1})\) and \(X' \sim \text{Bin}(n, \frac{\lambda}{n})\),
\[
\mathbb{P}(X \leq k) < \mathbb{P}(X' \leq k) \quad \text{for } k \leq \lambda - 1,
\]
and
\[
\mathbb{P}(X \leq k) > \mathbb{P}(X' \leq k) \quad \text{for } k \geq \lambda.
\]

Similarly, by taking \((p_1, \ldots, p_n) = (1, \frac{\lambda}{n - 1}, \ldots, \frac{\lambda}{n - 1})\),

\[
\mathbb{P}(X \leq k - 1) < \mathbb{P}(X' \leq k) \quad \text{for } k \leq \lambda - 1,
\]
and
\[
\mathbb{P}(X \leq k - 1) > \mathbb{P}(X' \leq k) \quad \text{for } k \geq \lambda.
\]

These inequalities were used in \([5]\) to derive the monotonicity of error in approximating the binomial distribution by a Poisson distribution. By letting \(X \sim \text{Bin}(n, p)\) and \(Y \sim \text{Poi}(np)\), they proved \(\mathbb{P}(X \leq k) - \mathbb{P}(Y \leq k)\) is positive if \(k \leq n^2 p/(n + 1)\) and is negative if \(k \geq np\). 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 \(\mu\) by at most 1. Precisely, he proved that
\[
m = \begin{cases} k & \text{if } k \leq \mu < k + \frac{1}{k+2}, \\ k + 1 & \text{if } k + 1 \leq \mu \leq k + 1 - \frac{1}{n-k+1}, \\ \mu & \text{if } k + 1 - \frac{1}{n-k+1} < \mu \leq k + 1. \end{cases}
\]

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 \(\xi_1, \ldots, \xi_n\) with \(0 \leq \xi_i \leq 1\),
\[
\mathbb{P}\left(\sum_{i=1}^{n} \xi_i \geq t\right) \leq \left(\frac{\mu}{t}\right)^t \left(\frac{n - \mu}{n-t}\right)^{n-t} \quad \text{for } t > \mu,
\]
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, \(\xi_i\) 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 \(\text{PB}(\mathbb{E}\xi_1, \ldots, \mathbb{E}\xi_n)\). By Theorem 2.1(2) with \(g(x) = x^k\), \(k \geq 1\), the moment generating function of \(\sum_{i=1}^{n} \xi_i\) is bounded by that of the binomial variable \(\text{Bin}(n, \mu/n)\). It then suffices to apply the Chernoff bound to get the Azuma–Hoeffding inequality (4).
To proceed further, we need some vocabularies. Let \( \{x_1, \ldots, x_n\} \) be the order statistics of \( \{x_1, \ldots, x_n\} \).

**Definition 1.** The vector \( \mathbf{x} \) is said to majorize the vector \( \mathbf{y} \), denoted as \( \mathbf{x} \succeq \mathbf{y} \), if

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

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

**Theorem 2.2.** Let \( X \sim \text{PB}(p_1, \ldots, p_n) \) and \( X' \sim \text{PB}(p_1', \ldots, p_n') \), and \( Y \sim \text{Bin}(n, p) \).

1. [49, 107] If \( (p_1, \ldots, p_n) \succeq (p_1', \ldots, p_n') \), then

\[
P(X \leq k) \leq P(X' \leq k) \quad \text{for } 0 \leq k \leq np^2 - 2,
\]

and

\[
P(X \leq k) \geq P(X' \leq k) \quad \text{for } np^2 + 2 \leq k \leq n.
\]

Moreover, \( \text{Var}(X) \leq \text{Var}(X') \).

2. [86] If \((−\log p_1, \ldots, −\log p_n) \succeq (−\log p_1', \ldots, −\log p_n')\), then \( X \) is stochastically larger than \( X' \), that is, \( P(X \geq k) \leq P(X' \geq k) \) for all \( k \).

3. [19] \( X \) is stochastically larger than \( Y \) if and only if \( p \leq (\prod_{i=1}^{n} p_i^{\frac{1}{n}}) \), and \( X \) is stochastically smaller than \( Y \) if and only if \( p \geq 1 − (\prod_{i=1}^{n} (1 − p_i^{\frac{1}{n}})) \). Consequently, if \( (\prod_{i=1}^{n} p_i^{\frac{1}{n}}) \geq 1 − (\prod_{i=1}^{n} (1 − p_i^{\frac{1}{n}})) \) then \( X \) is stochastically larger than \( X' \).

The proof of Theorem 2.2 relies on the fact that \( \mathbf{x} \succeq \mathbf{y} \) implies the components of \( \mathbf{x} \) are more spread out than those of \( \mathbf{y} \). For example in part (1), it boils down to proving if \( k \leq np^2 - 2 \), \( P(X \leq k) \) is a Schur concave function in \( p \), meaning its value increases as the components of \( p \) 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 \( X \sim \text{PB}(p_1, \ldots, p_n) \) and \( \mu := \sum_{i=1}^{n} p_i \). Then

\[
\frac{1}{32} \min\left(1, \frac{1}{\mu}\right) \sum_{i=1}^{n} p_i^2 \leq d_{\text{TV}}(\mathcal{L}(X), \text{Poi}(\mu))
\]

\[
(5)
\]

where \( d_{\text{TV}}(\cdot, \cdot) \) 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 \( \xi_1, \ldots, \xi_n \) independent Bernoulli random variables with parameters \( p_1, \ldots, p_n \),

\[
\mathbb{E}(\mu f(X + 1) - X f(X))
\]

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

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 \( \mu \), the upper bound in (5) is sharp. In particular, for \( \mu \leq 1 \), by taking \( p_1 = \mu \) and \( p_2 = \cdots = p_n = 0 \), we have

\[
d_{\text{TV}}(\mathcal{L}(X), \text{Poi}(\mu)) = \mu(1 - e^\mu) = \frac{1 - e^{-\mu}}{\mu} \sum_{i=1}^{n} p_i^2.
\]

- For large \( \mu \), 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/variance-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 \( \mu \) 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 \( (\mu, \sigma^2) \), denoted as \( \text{TP}(\mu, \sigma^2) \), if \( X - \mu + \sigma^2 + (\mu - \sigma^2) \sim \text{Poi}(\sigma^2 + (\mu - \sigma^2)) \), where \( \{\cdot\} \) is the fraction part of a positive number.
It is easy to see that a TP($\mu$, $\sigma^2$) random variable has mean $\mu$, and variance $\sigma^2 + (\mu + \sigma^2)$ which is between $\sigma^2$ 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 $X \sim \text{PB}(p_1, \ldots, p_n)$, and $\mu := \sum_{i=1}^n p_i$ and $\sigma^2 := \sum_{i=1}^n p_i(1 - p_i)$. Then

$$d_{TV}(\mathcal{L}(X), \text{TP}(\mu, \sigma^2)) \leq \frac{2 + \sqrt{\sum_{i=1}^n p_i^2(1 - p_i)}}{\sigma^2},$$

(6)

where $d_{TV}(\cdot, \cdot)$ 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 $k = 1, \ldots, n$, and units in each pair are indexed by $i = 1, 2$. The pair $k$ is matched on a set of observed covariates $x_{k1} = x_{k2}$, 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 $d_k \geq 0$ and $c_{ki} \in \{0, 1\}$. Here $c_{ki}$ represents the potential outcome which depends on the response $(r_{11}, r_{12}, \ldots, r_{n1}, r_{n2})$. 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

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

(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 $T \sim \text{PB}(p_1, \ldots, p_n)$. For $1 \leq k \leq n$, 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 $(\pi_1, \ldots, \pi_n)$ 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 $\mathcal{R}(t, \alpha) := \{(\pi_1, \ldots, \pi_n) : \mathbb{P}(T \geq t) \leq \alpha\}$, 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:

$$\max_{\pi \in C_\Gamma} \Gamma,$$

(8)

s.t. $\max_{\pi \in C_\Gamma} \mathbb{P}(T(\pi_1, \ldots, \pi_n) \geq t) \leq \alpha$,

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

$$\min_{A \in [0, \ldots, K]} \min_{\pi \in C_\Gamma} \sum_{k=0}^K \frac{1}{k!} \int_{-\infty}^x \phi(y) dy$$

(9)

s.t. $K = t - A$, $\lambda = \sum_{k=1}^n p_k - A$,

$$A \leq \sum_{k=1}^n p_k^2 < A + 1.$$

**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 $X \sim \text{PB}(p_1, \ldots, p_n)$, and $\mu := \sum_{i=1}^n p_i$ 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

$$\max_{0 \leq k \leq n} \left| \mathbb{P}(X = k) - \phi \left( \frac{k - \mu}{\sigma} \right) \right| \leq \frac{C}{\sigma},$$

(10)

2. [97] We have

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

(11)

Other than uniform bounds (10)–(11), several authors [16, 50, 88] studied error bounds for the normal approximation in other metrics. For $\mu$, $\nu$ 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
\[ W_r(\mu, \nu) := \inf\pi \left( \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x - y|^r \, \pi(dx \, dy) \right)^{\frac{1}{r}}, \]

where the infimum runs over all probability measures $\pi$ on $\mathbb{R} \times \mathbb{R}$ with marginals $\mu$ and $\nu$.

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

**Theorem 3.4.** Let $X \sim \text{PB}(p_1, \ldots, p_n)$, and $\mu := \sum_{i=1}^n p_i$ 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
\[ d_r(\mathcal{L}(X), N'(\mu, \sigma^2)) \leq \frac{C}{\sigma} \quad \text{for all } r \geq 1. \]

2. [16, 88] For each $r \geq 1$, there exists a constant $C_r > 0$ such that
\[ W_r(\mathcal{L}(X), N'(\mu, \sigma^2)) \leq \frac{C_r}{\sigma}. \]

Goldstein [50] proved $L'$ bound (12) for $r = 1$ with $C = 1$ via zero bias transformation. The general case follows from the inequality $d_r(\mu, \nu)^r \leq d_{r+1}(\mu, \nu)^{r-1}d_1(\mu, \nu)$ together with $L^1$ bound and the uniform bound (11). By the Kantorovich–Rubinstein duality, $d_1(\mu, \nu) = W_1(\mu, \nu)$. 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 $\xi_i$’s independent, $\mathbb{E}\xi_i = 0$ and $\sum_{i=1}^n \text{Var}(\xi_i) = 1$,

\[ W_r(\mathcal{L}(Z), N'(0, 1)) \leq C_r \left( \sum_{i=1}^n \mathbb{E}|Z_i|^{r+1} \right)^{\frac{1}{r}}. \]

This result was proved in [88] for $1 \leq r \leq 2$, and generalized to all $r \geq 1$ 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 $X \sim \text{PB}(p_1, \ldots, p_n)$,

\[ d_{TV}(\mathcal{L}(X), \text{Bin}(n, \mu/n)) \leq \frac{1 - (\mu/n)^{n+1} - (1 - \mu/n)^{n+1} \sum_{i=1}^n (p_i - \mu/n)^2}{(n + 1)(1 - \mu/n)\mu/n}. \]

Ehm’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 $X \sim \text{PB}(p_1, \ldots, p_n)$, and $\mu := \sum_{i=1}^n p_i$. For $m \geq 1$, let $d_m := d_{TV}(\mathcal{L}(X), \text{Bin}(m, \mu/m))$. Then for $m > \max\left\{ \frac{\mu^2}{2} \right\}$, the $\text{Bin}(m, \mu/m)$ approximation is strictly better than the Poisson approximation. It was also conjectured that the best $\text{Bin}(m, \mu/m)$ approximation is achieved for $m = \lfloor \frac{\mu^2}{2} \rfloor$ 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 $X \sim \text{PB}(p_1, \ldots, p_n)$, the probability generating function (PGF) of $X$ is

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

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 $(a_0, \ldots, a_n)$ be a sequence of nonnegative real numbers with $a_n > 0$. The associated generating polynomial is $f(z) := \sum_{i=0}^n a_iz^i$. Then the following conditions are equivalent:

1. The polynomial $f(z)$ has only real roots.
2. The sequence $(a_0/f(1), \ldots, a_n/f(1))$ is the probability distribution of a PB$(p_1, \ldots, p_n)$ distribution for some $p_i > 0$.
3. The sequence $(a_0, \ldots, a_n)$ is a Pólya frequency (PF) sequence. That is, the Toeplitz matrix $(a_{i+j})_{0 \leq i, j \leq n}$ (with convention $a_k = 0$ for $k < 0$) is totally nonnegative: every minor of $(a_{i+j})_{0 \leq i, j \leq n}$ 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 $(a_{i+j})_{0 \leq i, j \leq n}$ 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 $\text{HyperGeo}(n, K, N)$ 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 $X \sim \text{HyperGeo}(n, K, N)$, the probability mass function of $X$ is

$$P(X = k) = \binom{K}{k} \binom{N-K}{n-k} \binom{N}{n}^{-1} \quad \text{for } 0 \leq k \leq K.$$  \hspace{1cm} (17)

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

$$P(X = k) = \binom{n}{k} \binom{N-n}{K-k} \binom{N}{n}^{-1} \quad \text{for } 0 \leq k \leq K,$$

thus, the PGF of $X$ is

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

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

$$P_{K}^{a,b}(x) := \frac{1}{2^{K}} \sum_{k=0}^{K} \binom{K+a}{k} \binom{K+b}{k} \binom{K}{k} (x-1)^{K-k} (x+1)^{k}.$$

(19)

for $-1 < x < 1$, where $a, b > -1$. It is well known that for any fixed $a, b$, the family $(P_{K}^{a,b}, K \geq 0)$ 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

$$f(z) = \frac{\left| \frac{f(z)}{(z-1)^{K}} \right|_{z = 1+i}}{(N+1)} P_{K}^{a,b}(x).$$

(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 $X \sim \text{PB}(p_1, \ldots, p_n)$ for some $p_i$ if and only if $X$ is strong Rayleigh on $[0, \ldots, n]$.

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 $a_i \geq 0$ strong Rayleigh if it satisfies one of the conditions in Theorem 4.1.

For $n \geq 5$, 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:

$$a_{i}^{2} \geq a_{i-1}a_{i+1} \left(1 + \frac{1}{i}\right) \left(1 + \frac{1}{n-i}\right),$$

$$1 \leq i \leq n - 2.$$  \hspace{1cm} (21)

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

$$a_i^2 > 4a_{i-1}a_{i+1}, \quad 1 \leq i \leq n - 2.$$  \hspace{1cm} (22)

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

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

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 $\Lambda$, whose correlation functions at $(x_1, \ldots, x_n) \in \Lambda^{n}$ (i.e., the probability densities of having random points at $x_1, \ldots, x_n$) are

$$\rho(x_1, \ldots, x_n) = \det \left[K(x_i, x_j)\right]_{1 \leq i, j \leq n},$$

for some kernel $K : \Lambda \times \Lambda \to \mathbb{R}$. In the context of recommender systems, let $\Lambda$ 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 \( S_1^{(n)}, \ldots, S_d^{(n)} \) disjoint Borel sets indexed by \( n \geq 1 \), let \( X_{n,i} \) be the number of points of the DPP in \( S_i^{(n)} \), \( 1 \leq i \leq d \). It was proved in [99] that under fairly general conditions on the kernel \( K \) and \( S_1^{(n)}, \ldots, S_d^{(n)} \), the vector of counts \((X_{n,1}, \ldots, X_{n,d})\) has a multivariate Gaussian limit after suitable scaling. This is equivalent to the convergence in law:

\[
\frac{\sum_{i=1}^d \alpha_i X_{n,i} - \mathbb{E}(\sum_{i=1}^d \alpha_i X_{n,i})}{\text{Var}(\sum_{i=1}^d \alpha_i X_{n,i})} \xrightarrow{d} \mathcal{N}(0, 1),
\]

for any real numbers \( \alpha_1, \ldots, \alpha_d \). The proof of the convergence law (23) relies on the fact that the \( k \)th 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

\[
f(z_1, \ldots, z_d) := \mathbb{E} \prod_{i=1}^d z_i^{X_i}
\]

(24)

be the multivariate PGF of the counts \((X_1, \ldots, X_d)\) of a DPP. It is well known that if each \( X_i \) is bounded, \( f \) is real stable in the sense that if \( f = 0 \) has no solution \((z_1, \ldots, z_d)\) with each \( z_i \) having positive imaginary part. Similar to the univariate case, a nonnegative integer-valued random vector \((X_1, \ldots, X_d)\) 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 \((X_{n,1}, \ldots, X_{n,d}) \in \{0, \ldots, n\}^d\) a sequence of strong Rayleigh vectors, let \( \Sigma_n \) be its covariance matrix, and \( \sigma_n^2 \) be the \( \ell_2 \)-norm of \( \Sigma_n \). It was proved in [47] that under the conditions \( \sigma_n \to \infty \),

\[
\frac{\Sigma_n \sigma_n^{-2}}{\Sigma^*} \to \frac{\Sigma}{\Sigma^*} \quad \text{and} \quad \frac{\sigma_n}{n^{\frac{1}{2}}} \to \infty,
\]

(25)

\[
\frac{(X_{n,1}, \ldots, X_{n,d}) - \mathbb{E}(X_{n,1}, \ldots, X_{n,d})}{\sigma_n} \xrightarrow{d} \mathcal{N}(0, \Sigma^*),
\]

where \( \mathcal{N}(0, \Sigma^*) \) denotes the multivariate Gaussian distribution with mean 0 and covariance matrix \( \Sigma^* \). 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 \( q_1, \ldots, q_d \), 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 \((X_{n,1}, \ldots, X_{n,d})\). The condition \( \sigma_n/n^{\frac{1}{2}} \to \infty \) (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 \( q_i = 0 \), \( i \geq 2 \); if \( X \) is strong Rayleigh, how well can one approximate \( JX/k \) for each \( j, k \geq 1 \) 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 \( \lfloor X/k \rfloor \) is strong Rayleigh for each \( k \geq 1 \).

The key to the proof of Theorem 4.3 is that for a polynomial of degree \( n \) and \( k \geq 1 \), write \( f(z) = \sum_{j=0}^n x_j g_j(z^k) \), with \( g_j \) a polynomial of degree \( \lfloor n/j \rfloor \). 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 \( z_{n-k} < \cdots < z_{k} < 0 \), then the roots of \( g_i \) are \( z_i, \bar{z}_i+k, \bar{z}_i+2k \ldots \) In fact, the real-rootedness follows from the fact that (\( \alpha_n; n \geq 0 \)) is a PF sequence implies (\( \alpha_{kn+j}; n \geq 0 \)) is a PF sequence for each \( k \geq 1 \) and \( 0 < j < k \). This result is well known, see [2, 24]. But the root interlacing seems less obvious by PF sequences.

A natural question is whether \( \lfloor X/k \rfloor \) is strong Rayleigh for each \( j, k \geq 1 \). It turns out that \( \lfloor 2X/3 \rfloor \) can be far away from being strong Rayleigh. In fact, one can prove that for \( X \sim \text{Bin}(3n, 1/2) \), and \( z_i \) being the roots of the PGF of \( \lfloor 2X/3 \rfloor \), \( \max_i |\Im(z_i)| \geq 3^{2n-3n-2}/2 \) 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: \( \mathbb{P}(\lfloor 2X/3 \rfloor = 2k) = (3n+1)/(3_{k+1}) \) while \( \mathbb{P}(\lfloor 2X/3 \rfloor = 2k+1) = (3n+1)/(3_{k+1}) \). 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 \( \lfloor 2X/3 \rfloor \) 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 \( [2X/3] \) is Hurwitz stable. That is, all its roots have negative real parts.

The idea is to write the PGF of \( [2X/3] \) 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 \( [2X/3] \) can be factorized into polynomials with positive coefficients of degrees no greater than 2. Thus, \( [2X/3] \) 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 \( \{0, 1, \ldots, j\} \). 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 \( \hat{P} \) such that \( \mathbb{P}(d_{TV}(\hat{P}, P) \leq \epsilon) \geq 1 - \delta. \) The performance of a learning algorithm is measured by its sample complexity and its computational complexity.

For \( X \sim \text{PB}(p_1, \ldots, p_n) \), this amounts to finding a vector \( (\hat{p}_1, \ldots, \hat{p}_n) \) defining \( \hat{X} \sim \text{PB}(\hat{p}_1, \ldots, \hat{p}_n) \) such that \( d_{TV}(\hat{X}, X) \) 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 \( X \sim \text{PB}(p_1, \ldots, p_n) \) with unknown \( p_i \)'s. There is an algorithm such that given \( \epsilon, \delta > 0 \), it requires:

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

to construct a vector \( (\hat{p}_1, \ldots, \hat{p}_n) \) satisfying \( \mathbb{P}(d_{TV}(\hat{X}, X) \leq \epsilon) \geq 1 - \delta \) for \( \hat{X} \sim \text{PB}(\hat{p}_1, \ldots, \hat{p}_n) \).

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 Birge’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 Bin(\( n, 1/2 \)) from 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 \( X \sim \text{PB}(p_1, \ldots, p_n) \) from (1). Given \( p_1, \ldots, p_n \), 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 \( 0 \leq k \leq m \leq n \), let

\[
R_{k,m} := \mathbb{P}(X_m = k), \quad \text{where } X_m \sim \text{PB}(p_1, \ldots, p_m).
\]

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

- [45] For \( 0 \leq k \leq m \leq n \),
  \[
  R_{k,m} = (1 - p_m)R_{k,m-1} + p_mR_{k-1,m-1},
  \]
  with the convention that \( R_{-1,m} = R_{m+1,0} = 0 \) for \( 0 \leq m \leq n - 1 \), and \( R_{0,0} = 1 \).
- [30, 101] For \( i \geq 1 \), let \( t_i := \sum_{j=1}^{n} (p_j / (1 - p_j))^j \). Then
  \[
  R_{0,n} = \prod_{j=1}^{n} (1 - p_j),
  \]
  \[
  R_{k,n} = \frac{1}{k} \sum_{i=1}^{k} (-1)^{i+1} t_i R_{k-i,n}, \quad 1 \leq k \leq n.
  \]

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 \( R_{0,n} \), \ldots, \( R_{k-1,n} \). Moreover, it takes \( n - 1 \) additions to get \( t_1 \), and \( n - 1 \) additions and \( n \) multiplications to get \( t_i \) for \( i \geq 2 \). 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 \( \{R_{0,n}, \ldots, R_{n,n}\} \) (with memory being allowed), the recursion (26) requires approximately \( 3n^2/2 \) operations, and
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:

\begin{equation}
\mathbb{P}(X = k) = \frac{1}{n+1} \sum_{j=0}^{n} \exp(-i\omega j) x_j,
\end{equation}

and

\begin{equation}
\mathbb{P}(X \leq k) = \frac{1}{n+1} \sum_{j=0}^{n} \frac{1 - \exp(-i\omega (k+1) j)}{1 - \exp(-i\omega j)} x_j,
\end{equation}

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 \( \{x_0, \ldots, x_n\} \). Further, the r.h.s. of (28) is the discrete Fourier transform of \( \{x_0, \ldots, x_n\} \) which can be easily computed by Fast Fourier Transform in \( O(n \log n) \) 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-STSS852SUPP; .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**


