

## MIT Open Access Articles

### *Sparse covers for sums of indicators*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Daskalakis, Constantinos, and Christos Papadimitriou. "Sparse Covers for Sums of Indicators." *Probab. Theory Relat. Fields* 162, no. 3–4 (November 2, 2014): 679–705.

**As Published:** <http://dx.doi.org/10.1007/s00440-014-0582-8>

**Publisher:** Springer-Verlag

**Persistent URL:** <http://hdl.handle.net/1721.1/99971>

**Version:** Original manuscript: author's manuscript prior to formal peer review

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike



# Sparse Covers for Sums of Indicators

Constantinos Daskalakis\*  
EECS and CSAIL, MIT  
costis@mit.edu

Christos Papadimitriou†  
Computer Science, U.C. Berkeley  
christos@cs.berkeley.edu

October 3, 2014

## Abstract

For all  $n, \epsilon > 0$ , we show that the set of Poisson Binomial distributions on  $n$  variables admits a proper  $\epsilon$ -cover in total variation distance of size  $n^2 + n \cdot (1/\epsilon)^{O(\log^2(1/\epsilon))}$ , which can also be computed in polynomial time. We discuss the implications of our construction for approximation algorithms and the computation of approximate Nash equilibria in anonymous games.

## 1 Introduction

A *Poisson Binomial Distribution of order  $n$*  is the discrete probability distribution of the sum of  $n$  independent indicator random variables. The distribution is parameterized by a vector  $(p_i)_{i=1}^n \in [0, 1]^n$  of probabilities, and is denoted  $\text{PBD}(p_1, \dots, p_n)$ . In this paper we establish that the set  $\mathcal{S}_n$  of all Poisson Binomial distributions of order  $n$  admits certain useful covers with respect to the total variation distance  $d_{\text{TV}}(\cdot, \cdot)$  between distributions. Namely

**Theorem 1** (Main Theorem). *For all  $n, \epsilon > 0$ , there exists a set  $\mathcal{S}_{n, \epsilon} \subset \mathcal{S}_n$  such that:*

1.  $\mathcal{S}_{n, \epsilon}$  is an  $\epsilon$ -cover of  $\mathcal{S}_n$  in total variation distance; that is, for all  $D \in \mathcal{S}_n$ , there exists some  $D' \in \mathcal{S}_{n, \epsilon}$  such that  $d_{\text{TV}}(D, D') \leq \epsilon$
2.  $|\mathcal{S}_{n, \epsilon}| \leq n^2 + n \cdot \left(\frac{1}{\epsilon}\right)^{O(\log^2 1/\epsilon)}$
3.  $\mathcal{S}_{n, \epsilon}$  can be computed in time  $O(n^2 \log n) + O(n \log n) \cdot \left(\frac{1}{\epsilon}\right)^{O(\log^2 1/\epsilon)}$ .

Moreover, all distributions  $\text{PBD}(p_1, \dots, p_n) \in \mathcal{S}_{n, \epsilon}$  in the cover satisfy at least one of the following properties, for some positive integer  $k = k(\epsilon) = O(1/\epsilon)$ :

- (*k-sparse form*) there is some  $\ell \leq k^3$  such that, for all  $i \leq \ell$ ,  $p_i \in \left\{\frac{1}{k^2}, \frac{2}{k^2}, \dots, \frac{k^2-1}{k^2}\right\}$  and, for all  $i > \ell$ ,  $p_i \in \{0, 1\}$ ; or
- ( *$(n, k)$ -Binomial form*) there is some  $\ell \in \{1, \dots, n\}$  and  $q \in \left\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}\right\}$  such that, for all  $i \leq \ell$ ,  $p_i = q$  and, for all  $i > \ell$ ,  $p_i = 0$ ; moreover,  $\ell$  and  $q$  satisfy  $\ell q \geq k^2$  and  $\ell q(1 - q) \geq k^2 - k - 1$ .

---

\*Supported by a Sloan Foundation fellowship, a Microsoft Research faculty fellowship and NSF Award CCF-0953960 (CAREER) and CCF-1101491.

†Supported by NSF grant CCF-0964033 and a Google University Research Award.

Covers such as the one provided by Theorem 1 are of interest in the design of algorithms, when one is searching a class of distributions  $C$  to identify an element of the class with some quantitative property, or in optimizing over a class with respect to some objective. If the metric used in the construction of the cover is relevant for the problem at hand, and the cover is discrete, relatively small and easy to construct, then one can provide a useful approximation to the sought distribution by searching the cover, instead of searching all of  $C$ . For example, it is shown in [DP07, DP09, DP13] that Theorem 1 implies efficient algorithms for computing approximate Nash equilibria in an important class of multiplayer games, called anonymous [Mil96, Blo99].

We proceed with a fairly detailed sketch of the proof of our main cover theorem, Theorem 1, stating two additional results, Theorems 2 and 3. The complete proofs of Theorems 1, 2 and 3 are deferred to Sections 3, 4 and 5 respectively. Section 1.4 discusses related work, while Section 2 provides formal definitions, as well as known approximations to the Poisson Binomial distribution by simpler distributions, which are used in the proof.

## 1.1 Proof Outline and Additional Results

At a high level, the proof of Theorem 1 is obtained in two steps. First, we establish the existence of an  $\epsilon$ -cover whose size is polynomial in  $n$  and  $(1/\epsilon)^{1/\epsilon^2}$ , via Theorem 2. We then show that this cover can be pruned to size polynomial in  $n$  and  $(1/\epsilon)^{\log^2(1/\epsilon)}$  using Theorem 3, which provides a quantification of how the total variation distance between Poisson Binomial distributions depends on the number of their first moments that are equal.

We proceed to state the two ingredients of the proof, Theorems 2 and 3. We start with Theorem 2 whose detailed sketch is given in Section 1.2, and complete proof in Section 4.

**Theorem 2.** *Let  $X_1, \dots, X_n$  be arbitrary mutually independent indicators, and  $k \in \mathbb{N}$ . Then there exist mutually independent indicators  $Y_1, \dots, Y_n$  satisfying the following:*

1.  $d_{\text{TV}}(\sum_i X_i, \sum_i Y_i) \leq 41/k$ ;
2. *at least one of the following is true:*
  - (a) *( $k$ -sparse form) there exists some  $\ell \leq k^3$  such that, for all  $i \leq \ell$ ,  $\mathbb{E}[Y_i] \in \left\{ \frac{1}{k^2}, \frac{2}{k^2}, \dots, \frac{k^2-1}{k^2} \right\}$  and, for all  $i > \ell$ ,  $\mathbb{E}[Y_i] \in \{0, 1\}$ ; or*
  - (b) *(( $n, k$ )-Binomial form) there is some  $\ell \in \{1, \dots, n\}$  and  $q \in \left\{ \frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n} \right\}$  such that, for all  $i \leq \ell$ ,  $\mathbb{E}[Y_i] = q$  and, for all  $i > \ell$ ,  $\mathbb{E}[Y_i] = 0$ ; moreover,  $\ell$  and  $q$  satisfy  $\ell q \geq k^2$  and  $\ell q(1-q) \geq k^2 - k - 1$ .*

Theorem 2 implies the existence of an  $\epsilon$ -cover of  $\mathcal{S}_n$  whose size is  $n^2 + n \cdot (1/\epsilon)^{O(1/\epsilon^2)}$ . This cover can be obtained by enumerating over all Poisson Binomial distributions of order  $n$  that are in  $k$ -sparse or  $(n, k)$ -Binomial form as defined in the statement of the theorem, for  $k = \lceil 41/\epsilon \rceil$ .

The next step is to sparsify this cover by removing elements to obtain Theorem 1. Note that the term  $n \cdot (1/\epsilon)^{O(1/\epsilon^2)}$  in the size of the cover is due to the enumeration over distributions in sparse form. Using Theorem 3 below, we argue that there is a lot of redundancy in those distributions, and that it suffices to only include  $n \cdot (1/\epsilon)^{O(\log^2 1/\epsilon)}$  of them in the cover. In particular, Theorem 3 establishes that, if two Poisson Binomial distributions have their first  $O(\log 1/\epsilon)$  moments equal, then their distance is at most  $\epsilon$ . So we only need to include at most one sparse form distribution with the same first  $O(\log 1/\epsilon)$  moments in our cover. We proceed to state Theorem 3, postponing its proof to Section 5. In Section 1.3 we provide a sketch of the proof.

**Theorem 3.** Let  $\mathcal{P} := (p_i)_{i=1}^n \in [0, 1/2]^n$  and  $\mathcal{Q} := (q_i)_{i=1}^n \in [0, 1/2]^n$  be two collections of probability values. Let also  $\mathcal{X} := (X_i)_{i=1}^n$  and  $\mathcal{Y} := (Y_i)_{i=1}^n$  be two collections of mutually independent indicators with  $\mathbb{E}[X_i] = p_i$  and  $\mathbb{E}[Y_i] = q_i$ , for all  $i \in [n]$ . If for some  $d \in [n]$  the following condition is satisfied:

$$(C_d) : \quad \sum_{i=1}^n p_i^\ell = \sum_{i=1}^n q_i^\ell, \quad \text{for all } \ell = 1, \dots, d,$$

$$\text{then } d_{\text{TV}} \left( \sum_i X_i, \sum_i Y_i \right) \leq 13(d+1)^{1/4} 2^{-(d+1)/2}. \quad (1)$$

**Remark 1.** Condition  $(C_d)$  in the statement of Theorem 3 constrains the first  $d$  power sums of the expectations of the constituent indicators of two Poisson Binomial distributions. To relate these power sums to the moments of these distributions we can use the theory of symmetric polynomials to arrive at the following equivalent condition to  $(C_d)$ :

$$(V_d) : \quad \mathbb{E} \left[ \left( \sum_{i=1}^n X_i \right)^\ell \right] = \mathbb{E} \left[ \left( \sum_{i=1}^n Y_i \right)^\ell \right], \quad \text{for all } \ell \in [d].$$

We provide a proof that  $(C_d) \Leftrightarrow (V_d)$  in Proposition 2 of Section 6.

**Remark 2.** In view of Remark 1, Theorem 3 says the following:

“If two sums of independent indicators with expectations in  $[0, 1/2]$  have equal first  $d$  moments, then their total variation distance is  $2^{-\Omega(d)}$ .”

We note that the bound (1) does not depend on the number of variables  $n$ , and in particular does not rely on summing a large number of variables. We also note that, since we impose no constraint on the expectations of the indicators, we also impose no constraint on the variance of the resulting Poisson Binomial distributions. Hence we cannot use Berry-Esséen type bounds to bound the total variation distance of the two Poisson Binomial distributions by approximating them with Normal distributions. Finally, it is easy to see that Theorem 3 holds if we replace  $[0, 1/2]$  with  $[1/2, 1]$ . See Corollary 1 in Section 6.

In Section 3 we show how to use Theorems 2 and 3 to obtain Theorem 1. We continue with the outlines of the proofs of Theorems 2 and 3, postponing their complete proofs to Sections 4 and 5.

## 1.2 Outline of Proof of Theorem 2

Given arbitrary indicators  $X_1, \dots, X_n$  we obtain indicators  $Y_1, \dots, Y_n$ , satisfying the requirements of Theorem 2, in two steps. We first massage the given variables  $X_1, \dots, X_n$  to obtain variables  $Z_1, \dots, Z_n$  such that

$$d_{\text{TV}} \left( \sum_i X_i, \sum_i Z_i \right) \leq 7/k; \quad (2)$$

and  $\mathbb{E}[Z_i] \notin \left(0, \frac{1}{k}\right) \cup \left(1 - \frac{1}{k}, 1\right);$

that is, we eliminate from our collection variables that have expectations very close to 0 or 1, without traveling too much distance from the starting Poisson Binomial distribution.

Variables  $Z_1, \dots, Z_n$  do not necessarily satisfy Properties 2a or 2b in the statement of Theorem 2, but allow us to define variables  $Y_1, \dots, Y_n$  which do satisfy one of these properties and, moreover,

$$d_{\text{TV}} \left( \sum_i Z_i, \sum_i Y_i \right) \leq 34/k. \quad (3)$$

(2), (3) and the triangle inequality imply  $d_{\text{TV}}(\sum_i X_i, \sum_i Y_i) \leq \frac{41}{k}$ , concluding the proof of Theorem 2.

Let us call *Stage 1* the process of determining the  $Z_i$ 's and *Stage 2* the process of determining the  $Y_i$ 's. The two stages are described briefly below, and in detail in Sections 4.1 and 4.2 respectively. For convenience, we use the following notation: for  $i = 1, \dots, n$ ,  $p_i = \mathbb{E}[X_i]$  will denote the expectation of the given indicator  $X_i$ ,  $p'_i = \mathbb{E}[Z_i]$  the expectation of the intermediate indicator  $Z_i$ , and  $q_i = \mathbb{E}[Y_i]$  the expectation of the final indicator  $Y_i$ .

**Stage 1:** Recall that our goal in this stage is to define a Poisson Binomial distribution  $\sum_i Z_i$  whose constituent indicators have no expectation in  $\mathcal{T}_k := (0, \frac{1}{k}) \cup (1 - \frac{1}{k}, 1)$ . The expectations  $(p'_i = \mathbb{E}[Z_i])_i$  are defined in terms of the corresponding  $(p_i)_i$  as follows. For all  $i$ , if  $p_i \notin \mathcal{T}_k$  we set  $p'_i = p_i$ . Then, if  $\mathcal{L}_k$  is the set of indices  $i$  such that  $p_i \in (0, 1/k)$ , we choose any collection  $(p'_i)_{i \in \mathcal{L}_k}$  so as to satisfy  $|\sum_{i \in \mathcal{L}_k} p_i - \sum_{i \in \mathcal{L}_k} p'_i| \leq 1/k$  and  $p'_i \in \{0, 1/k\}$ , for all  $i \in \mathcal{L}_k$ . That is, we round all indicators' expectations to 0 or  $1/k$  while preserving the expectation of their sum, to within  $1/k$ . Using the Poisson approximation to the Poisson Binomial distribution, given as Theorem 4 in Section 2.1, we can argue that  $\sum_{i \in \mathcal{L}_k} X_i$  is within  $1/k$  of a Poisson distribution with the same mean. By the same token,  $\sum_{i \in \mathcal{L}_k} Z_i$  is  $1/k$ -close to a Poisson distribution with the same mean. And the two resulting Poisson distributions have means that are within  $1/k$ , and are therefore  $1.5/k$ -close to each other (see Lemma 3). Hence, by triangle inequality  $\sum_{i \in \mathcal{L}_k} X_i$  is  $3.5/k$ -close to  $\sum_{i \in \mathcal{L}_k} Z_i$ . A similar construction is used to define the  $p'_i$ 's corresponding to the  $p_i$ 's lying in  $(1 - 1/k, 1)$ . The details of this step can be found in Section 4.1.

**Stage 2:** The definition of  $(q_i)_i$  depends on the number  $m$  of  $p'_i$ 's which are not 0 or 1. The case  $m \leq k^3$  corresponds to Case 2a in the statement of Theorem 2, while the case  $m > k^3$  corresponds to Case 2b.

- Case  $m \leq k^3$ : First, we set  $q_i = p'_i$ , if  $p'_i \in \{0, 1\}$ . We then argue that each  $p'_i$ ,  $i \in \mathcal{M} := \{i \mid p'_i \notin \{0, 1\}\}$ , can be rounded to some  $q_i$ , which is an integer multiple of  $1/k^2$ , so that (3) holds. Notice that, if we were allowed to use multiples of  $1/k^4$ , this would be immediate via an application of Lemma 2:

$$d_{\text{TV}} \left( \sum_i Z_i, \sum_i Y_i \right) \leq \sum_{i \in \mathcal{M}} |p'_i - q_i|.$$

We improve the required accuracy to  $1/k^2$  via a series of Binomial approximations to the Poisson Binomial distribution, using Ehm's bound [Ehm91] stated as Theorem 5 in Section 2.1. The details involve partitioning the interval  $[1/k, 1 - 1/k]$  into irregularly sized subintervals, whose endpoints are integer multiples of  $1/k^2$ . We then round all but one of the  $p'_i$ 's falling in each subinterval to the endpoints of the subinterval so as to maintain their total expectation, and apply Ehm's approximation to argue that the distribution of their sum is not affected by more than  $O(1/k^2)$  in total variation distance. It is crucial that the total number of subintervals is  $O(k)$  to get a total hit of at most  $O(1/k)$  in variation distance in the overall distribution. The details are given in Section 4.2.1.

- Case  $m > k^3$ : We approximate  $\sum_i Z_i$  with a Translated Poisson distribution (defined formally in Section 2), using Theorem 6 of Section 2.1 due to Röllin [Rö07]. The quality of the approximation is inverse proportional to the standard deviation of  $\sum_i Z_i$ , which is at least  $k$ , by the assumption  $m > k^3$ . Hence, we show that  $\sum_i Z_i$  is  $3/k$ -close to a Translated Poisson distribution. We then argue that the latter is  $6/k$ -close to a Binomial distribution  $B(m', q)$ , where  $m' \leq n$  and  $q$  is an integer multiple of  $\frac{1}{n}$ . In particular, we show that an appropriate choice of  $m'$  and  $q$  implies (3), if we set  $m'$  of the  $q_i$ 's equal to  $q$  and the remaining equal to 0. The details are in Section 4.2.2.

### 1.3 Outline of Proof of Theorem 3

Using Roos's expansion [Roo00], given as Theorem 7 of Section 2.1, we express  $\text{PBD}(p_1, \dots, p_n)$  as a weighted sum of the Binomial distribution  $\mathcal{B}(n, p)$  at  $p = \bar{p} = \sum p_i/n$  and its first  $n$  derivatives with respect to  $p$  also at value  $p = \bar{p}$ . (These derivatives correspond to finite signed measures.) We notice that the coefficients of the first  $d+1$  terms of this expansion are symmetric polynomials in  $p_1, \dots, p_n$  of degree at most  $d$ . Hence, from the theory of symmetric polynomials, each of these coefficients can be written as a function of the power-sum symmetric polynomials  $\sum_i p_i^\ell$  for  $\ell = 1, \dots, d$ . So, whenever two Poisson Binomial distributions satisfy Condition  $(C_d)$ , the first  $d+1$  terms of their expansions are exactly identical, and the total variation distance of the distributions depends only on the other terms of the expansion (those corresponding to higher derivatives of the Binomial distribution). The proof is concluded by showing that the joint contribution of these terms to the total variation distance can be bounded by  $2^{-\Omega(d)}$ , using Proposition 1 of Section 2.1, which is also due to Roos [Roo00]. The details are provided in Section 5.

### 1.4 Related Work

It is believed that Poisson [Poi37] was the first to study the Poisson Binomial distribution, hence its name. Sometimes the distribution is also referred to as ‘‘Poisson’s Binomial Distribution.’’ PBDs have many uses in research areas such as survey sampling, case-control studies, and survival analysis; see e.g. [CL97] for a survey of their uses. They are also very important in the design of randomized algorithms [MR95].

In Probability and Statistics there is a broad literature studying various properties of these distributions; see [Wan93] for an introduction to some of this work. Many results provide approximations to the Poisson Binomial distribution via simpler distributions. In a well-known result, Le Cam [LC60] shows that, for any vector  $(p_i)_{i=1}^n \in [0, 1]^n$ ,

$$d_{\text{TV}} \left( \text{PBD}(p_1, \dots, p_n), \text{Poisson} \left( \sum_{i=1}^n p_i \right) \right) \leq \sum_{i=1}^n p_i^2,$$

where  $\text{Poisson}(\lambda)$  is the Poisson distribution with parameter  $\lambda$ . Subsequently many other proofs of this bound and improved ones, such as Theorem 4 of Section 2.1, were given, using a range of different techniques; [HC60, Che74, BH84, DP86] is a sampling of work along these lines, and Steele [Ste94] gives an extensive list of relevant references. Much work has also been done on approximating PBDs by Normal distributions (see e.g. [Ber41, Ess42, Mik93, Vol95, CGS10]) and by Binomial distributions; see e.g. Ehm’s result [Ehm91], given as Theorem 5 of Section 2.1, as well as Soon’s result [Soo96] and Roos’s result [Roo00], given as Theorem 7 of Section 2.1.

These results provide structural information about PBDs that can be well approximated by simpler distributions, but fall short of our goal of approximating a PBD to within *arbitrary accuracy*. Indeed, the approximations obtained in the probability literature (such as the Poisson, Normal and

Binomial approximations) typically depend on the first few moments of the PBD being approximated, while higher moments are crucial for arbitrary approximation [Roo00]. At the same time, algorithmic applications often require that the approximating distribution is of the same kind as the distribution that is being approximated. E.g., in the anonymous game application mentioned earlier, the parameters of the given PBD correspond to mixed strategies of players at Nash equilibrium, and the parameters of the approximating PBD correspond to mixed strategies at approximate Nash equilibrium. Approximating the given PBD via a Poisson or a Normal distribution would not have any meaning in the context of a game.

As outlined above, the proof of our main result, Theorem 1, builds on Theorems 2 and 3. A weaker form of these theorems was announced in [Das08, DP09], while a weaker form of Theorem 1 was announced in [DDS12].

## 2 Preliminaries

For a positive integer  $\ell$ , we denote by  $[\ell]$  the set  $\{1, \dots, \ell\}$ . For a random variable  $X$ , we denote by  $\mathcal{L}(X)$  its distribution. We further need the following definitions.

*Total variation distance:* For two distributions  $\mathbb{P}$  and  $\mathbb{Q}$  supported on a finite set  $A$  their *total variation distance* is defined as

$$d_{\text{TV}}(\mathbb{P}, \mathbb{Q}) := \frac{1}{2} \sum_{\alpha \in A} |\mathbb{P}(\alpha) - \mathbb{Q}(\alpha)|.$$

An equivalent way to define  $d_{\text{TV}}(\mathbb{P}, \mathbb{Q})$  is to view  $\mathbb{P}$  and  $\mathbb{Q}$  as vectors in  $\mathbb{R}^A$ , and define  $d_{\text{TV}}(\mathbb{P}, \mathbb{Q}) = \frac{1}{2} \|\mathbb{P} - \mathbb{Q}\|_1$  to equal half of their  $\ell_1$  distance. If  $X$  and  $Y$  are random variables ranging over a finite set, their total variation distance, denoted  $d_{\text{TV}}(X, Y)$ , is defined to equal  $d_{\text{TV}}(\mathcal{L}(X), \mathcal{L}(Y))$ .

*Covers:* Let  $\mathcal{F}$  be a set of probability distributions. A subset  $\mathcal{G} \subseteq \mathcal{F}$  is called a (*proper*)  $\epsilon$ -*cover* of  $\mathcal{F}$  in total variation distance if, for all  $D \in \mathcal{F}$ , there exists some  $D' \in \mathcal{G}$  such that  $d_{\text{TV}}(D, D') \leq \epsilon$ .

*Poisson Binomial Distribution:* A *Poisson Binomial distribution of order  $n \in \mathbb{N}$*  is the discrete probability distribution of the sum  $\sum_{i=1}^n X_i$  of  $n$  mutually independent Bernoulli random variables  $X_1, \dots, X_n$ . We denote the set of all Poisson Binomial distributions of order  $n$  by  $\mathcal{S}_n$ .

By definition, a Poisson Binomial distribution  $D \in \mathcal{S}_n$  can be represented by a vector  $(p_i)_{i=1}^n \in [0, 1]^n$  of probabilities as follows. We map  $D \in \mathcal{S}_n$  to a vector of probabilities by finding a collection  $X_1, \dots, X_n$  of mutually independent indicators such that  $\sum_{i=1}^n X_i$  is distributed according to  $D$ , and setting  $p_i = \mathbb{E}[X_i]$  for all  $i$ . The following lemma implies that the resulting vector of probabilities is unique up to a permutation, so that there is a one-to-one correspondence between Poisson Binomial distributions and vectors  $(p_i)_{i=1}^n \in [0, 1]^n$  such that  $0 \leq p_1 \leq p_2 \leq \dots \leq p_n \leq 1$ . The proof of this lemma can be found in Section 6.

**Lemma 1.** *Let  $X_1, \dots, X_n$  be mutually independent indicators with expectations  $p_1 \leq p_2 \leq \dots \leq p_n$  respectively. Similarly let  $Y_1, \dots, Y_n$  be mutually independent indicators with expectations  $q_1 \leq \dots \leq q_n$  respectively. The distributions of  $\sum_i X_i$  and  $\sum_i Y_i$  are different if and only if  $(p_1, \dots, p_n) \neq (q_1, \dots, q_n)$ .*

We will be denoting a Poisson Binomial distribution  $D \in \mathcal{S}_n$  by  $\text{PBD}(p_1, \dots, p_n)$  when it is the distribution of the sum  $\sum_{i=1}^n X_i$  of mutually independent indicators  $X_1, \dots, X_n$  with expectations  $p_i = \mathbb{E}[X_i]$ , for all  $i$ . Given the above discussion, the representation is unique up to a permutation of the  $p_i$ 's.

*Translated Poisson Distribution:* We say that an integer random variable  $Y$  has a *translated Poisson distribution* with parameters  $\mu$  and  $\sigma^2$  and write  $\mathcal{L}(Y) = TP(\mu, \sigma^2)$  iff

$$\mathcal{L}(Y - \lfloor \mu - \sigma^2 \rfloor) = \text{Poisson}(\sigma^2 + \{\mu - \sigma^2\}),$$

where  $\{\mu - \sigma^2\}$  represents the fractional part of  $\mu - \sigma^2$ .

*Order Notation:* Let  $f(x)$  and  $g(x)$  be two positive functions defined on some infinite subset of  $\mathbb{R}_+$ . One writes  $f(x) = O(g(x))$  if and only if, for sufficiently large values of  $x$ ,  $f(x)$  is at most a constant times  $g(x)$ . That is,  $f(x) = O(g(x))$  if and only if there exist positive real numbers  $M$  and  $x_0$  such that

$$f(x) \leq Mg(x), \text{ for all } x > x_0.$$

Similarly, we write  $f(x) = \Omega(g(x))$  if and only if there exist positive reals  $M$  and  $x_0$  such that

$$f(x) \geq Mg(x), \text{ for all } x > x_0.$$

We are casual in our use of the order notation  $O(\cdot)$  and  $\Omega(\cdot)$  throughout the paper. Whenever we write  $O(f(n))$  or  $\Omega(f(n))$  in some bound where  $n$  ranges over the integers, we mean that there exists a constant  $c > 0$  such that the bound holds true for sufficiently large  $n$  if we replace the  $O(f(n))$  or  $\Omega(f(n))$  in the bound by  $c \cdot f(n)$ . On the other hand, whenever we write  $O(f(1/\epsilon))$  or  $\Omega(f(1/\epsilon))$  in some bound where  $\epsilon$  ranges over the positive reals, we mean that there exists a constant  $c > 0$  such that the bound holds true for sufficiently *small*  $\epsilon$  if we replace the  $O(f(1/\epsilon))$  or  $\Omega(f(1/\epsilon))$  in the bound with  $c \cdot f(1/\epsilon)$ .

We conclude with an easy but useful lemma whose proof we defer to Section 6.

**Lemma 2.** *Let  $X_1, \dots, X_n$  be mutually independent random variables, and let  $Y_1, \dots, Y_n$  be mutually independent random variables. Then*

$$d_{\text{TV}} \left( \sum_{i=1}^n X_i, \sum_{i=1}^n Y_i \right) \leq \sum_{i=1}^n d_{\text{TV}}(X_i, Y_i).$$

## 2.1 Approximations to the Poisson Binomial Distribution

We present a collection of known approximations to the Poisson Binomial distribution via simpler distributions. The quality of these approximations can be quantified in terms of the first few moments of the Poisson Binomial distribution that is being approximated. We will make use of these bounds to approximate Poisson Binomial distributions in different regimes of their moments. Theorems 4–6 are obtained via the Stein-Chen method.

**Theorem 4** (Poisson Approximation [BH84, BHJ92]). *Let  $J_1, \dots, J_n$  be mutually independent indicators with  $\mathbb{E}[J_i] = t_i$ . Then*

$$d_{\text{TV}} \left( \sum_{i=1}^n J_i, \text{Poisson} \left( \sum_{i=1}^n t_i \right) \right) \leq \frac{\sum_{i=1}^n t_i^2}{\sum_{i=1}^n t_i}.$$

**Theorem 5** (Binomial Approximation [Ehm91]). *Let  $J_1, \dots, J_n$  be mutually independent indicators with  $\mathbb{E}[J_i] = t_i$ , and  $\bar{t} = \frac{\sum_{i=1}^n t_i}{n}$ . Then*

$$d_{\text{TV}} \left( \sum_{i=1}^n J_i, \mathcal{B}(n, \bar{t}) \right) \leq \frac{\sum_{i=1}^n (t_i - \bar{t})^2}{(n+1)\bar{t}(1-\bar{t})},$$

where  $\mathcal{B}(n, \bar{t})$  is the Binomial distribution with parameters  $n$  and  $\bar{t}$ .

**Theorem 6** (Translated Poisson Approximation [Rö7]). *Let  $J_1, \dots, J_n$  be mutually independent indicators with  $\mathbb{E}[J_i] = t_i$ . Then*

$$d_{\text{TV}} \left( \sum_{i=1}^n J_i, TP(\mu, \sigma^2) \right) \leq \frac{\sqrt{\sum_{i=1}^n t_i^3(1-t_i) + 2}}{\sum_{i=1}^n t_i(1-t_i)},$$

where  $\mu = \sum_{i=1}^n t_i$  and  $\sigma^2 = \sum_{i=1}^n t_i(1-t_i)$ .

The approximation theorems stated above do not always provide tight enough approximations. When these fail, we employ the following theorem of Roos [Roo00], which provides an expansion of the Poisson Binomial distribution as a weighted sum of a finite number of signed measures: the Binomial distribution  $\mathcal{B}(n, p)$  (for an arbitrary value of  $p$ ) and its first  $n$  derivatives with respect to the parameter  $p$ , at the chosen value of  $p$ . For the purposes of the following statement we denote by  $\mathcal{B}_{n,p}(m)$  the probability assigned by the Binomial distribution  $\mathcal{B}(n, p)$  to integer  $m$ .

**Theorem 7** ([Roo00]). *Let  $\mathcal{P} := (p_i)_{i=1}^n \in [0, 1]^n$ ,  $X_1, \dots, X_n$  be mutually independent indicators with expectations  $p_1, \dots, p_n$ , and  $X = \sum_i X_i$ . Then, for all  $m \in \{0, \dots, n\}$  and  $p \in [0, 1]$ ,*

$$\Pr[X = m] = \sum_{\ell=0}^n \alpha_\ell(\mathcal{P}, p) \cdot \delta^\ell \mathcal{B}_{n,p}(m), \quad (4)$$

where for the purposes of the above expression:

- $\alpha_0(\mathcal{P}, p) := 1$  and for  $\ell \in [n]$ :

$$\alpha_\ell(\mathcal{P}, p) := \sum_{1 \leq k(1) < \dots < k(\ell) \leq n} \prod_{r=1}^{\ell} (p_{k(r)} - p);$$

- and for all  $\ell \in \{0, \dots, n\}$ :

$$\delta^\ell \mathcal{B}_{n,p}(m) := \frac{(n-\ell)!}{n!} \frac{d^\ell}{dp^\ell} \mathcal{B}_{n,p}(m),$$

where for the last definition we interpret  $\mathcal{B}_{n,p}(m) \equiv \binom{n}{m} p^m (1-p)^{n-m}$  as a function of  $p$ .

We can use Theorem 7 to get tighter approximations to the Poisson Binomial distribution by appropriately tuning the number of terms of summation (4) that we keep. The following proposition, shown in the proof of Theorem 2 of [Roo00], bounds the  $\ell_1$  approximation error to the Poisson Binomial distribution when only the first  $d+1$  terms of summation (4) are kept. The error decays exponentially in  $d$  as long as the quantity  $\theta(\mathcal{P}, p)$  in the proposition statement is smaller than 1.

**Proposition 1** ([Roo00]). *Let  $\mathcal{P} = (p_i)_{i=1}^n \in [0, 1]^n$ ,  $p \in [0, 1]$ ,  $\alpha_\ell(\cdot, \cdot)$  and  $\delta^\ell \mathcal{B}_{n,p}(\cdot)$  as in the statement of Theorem 7, and take*

$$\theta(\mathcal{P}, p) = \frac{2 \sum_{i=1}^n (p_i - p)^2 + (\sum_{i=1}^n (p_i - p))^2}{2np(1-p)}.$$

If  $\theta(\mathcal{P}, p) < 1$ , then, for all  $d \geq 0$ :

$$\sum_{\ell=d+1}^n |\alpha_\ell(\mathcal{P}, p)| \cdot \|\delta^\ell \mathcal{B}_{n,p}(\cdot)\|_1 \leq \sqrt{e} (d+1)^{1/4} \theta(\mathcal{P}, p)^{(d+1)/2} \frac{1 - \frac{d}{d+1} \sqrt{\theta(\mathcal{P}, p)}}{(1 - \sqrt{\theta(\mathcal{P}, p)})^2},$$

where  $\|\delta^\ell \mathcal{B}_{n,p}(\cdot)\|_1 := \sum_{m=0}^n |\delta^\ell \mathcal{B}_{n,p}(m)|$ .

### 3 Proof of Theorem 1

We first argue that Theorem 2 already implies the existence of an  $\epsilon$ -cover  $\mathcal{S}'_{n,\epsilon}$  of  $\mathcal{S}_n$  of size at most  $n^2 + n \cdot \left(\frac{1}{\epsilon}\right)^{O(1/\epsilon^2)}$ . This cover is obtained by taking the union of all Poisson Binomial distributions in  $(n, k)$ -Binomial form and all Poisson Binomial distributions in  $k$ -sparse form, for  $k = \lceil 41/\epsilon \rceil$ . The total number of Poisson Binomial distributions in  $(n, k)$ -Binomial form is at most  $n^2$ , since there are at most  $n$  choices for the value of  $\ell$  and at most  $n$  choices for the value of  $q$ . The total number of Poisson Binomial distributions in  $k$ -sparse form is at most  $(k^3 + 1) \cdot k^{3k^2} \cdot (n + 1) = n \cdot \left(\frac{1}{\epsilon}\right)^{O(1/\epsilon^2)}$  since there are  $k^3 + 1$  choices for  $\ell$ , at most  $k^{3k^2}$  choices of probabilities  $p_1 \leq p_2 \leq \dots \leq p_\ell$  in  $\left\{\frac{1}{k^2}, \frac{2}{k^2}, \dots, \frac{k^2-1}{k^2}\right\}$ , and at most  $n + 1$  choices for the number of variables indexed by  $i > \ell$  that have expectation equal to 1.<sup>1</sup> Notice that enumerating over the above distributions takes time  $O(n^2 \log n) + O(n \log n) \cdot \left(\frac{1}{\epsilon}\right)^{O(1/\epsilon^2)}$ , as a number in  $\{0, \dots, n\}$  and a probability in  $\left\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}\right\}$  can be represented using  $O(\log n)$  bits, while a number in  $\{0, \dots, k^3\}$  and a probability in  $\left\{\frac{1}{k^2}, \frac{2}{k^2}, \dots, \frac{k^2-1}{k^2}\right\}$  can be represented using  $O(\log k) = O(\log 1/\epsilon)$  bits.

We next show that we can remove from  $\mathcal{S}'_{n,\epsilon}$  a large number of the sparse-form distributions it contains to obtain a  $2\epsilon$ -cover of  $\mathcal{S}_n$ . In particular, we shall only keep  $n \cdot \left(\frac{1}{\epsilon}\right)^{O(\log^2 1/\epsilon)}$  sparse-form distributions by appealing to Theorem 3. To explain the pruning we introduce some notation. For a collection  $\mathcal{P} = (p_i)_{i \in [n]} \in [0, 1]^n$  of probability values we denote by  $\mathcal{L}_{\mathcal{P}} = \{i \mid p_i \in (0, 1/2]\}$  and by  $\mathcal{R}_{\mathcal{P}} = \{i \mid p_i \in (1/2, 1)\}$ . Theorem 3, Corollary 1, Lemma 2 and Lemma 1 imply that if two collections  $\mathcal{P} = (p_i)_{i \in [n]}$  and  $\mathcal{Q} = (q_i)_{i \in [n]}$  of probability values satisfy

$$\begin{aligned} \sum_{i \in \mathcal{L}_{\mathcal{P}}} p_i^t &= \sum_{i \in \mathcal{L}_{\mathcal{Q}}} q_i^t, \quad \text{for all } t = 1, \dots, d; \\ \sum_{i \in \mathcal{R}_{\mathcal{P}}} p_i^t &= \sum_{i \in \mathcal{R}_{\mathcal{Q}}} q_i^t, \quad \text{for all } t = 1, \dots, d; \text{ and} \end{aligned}$$

$(p_i)_{[n] \setminus (\mathcal{L}_{\mathcal{P}} \cup \mathcal{R}_{\mathcal{P}})}$  and  $(q_i)_{[n] \setminus (\mathcal{L}_{\mathcal{Q}} \cup \mathcal{R}_{\mathcal{Q}})}$  are equal up to a permutation;

then  $d_{\text{TV}}(\text{PBD}(\mathcal{P}), \text{PBD}(\mathcal{Q})) \leq 2 \cdot 13(d+1)^{1/4} 2^{-(d+1)/2}$ . In particular, for some  $d(\epsilon) = O(\log 1/\epsilon)$ , this bound becomes at most  $\epsilon$ .

For a collection  $\mathcal{P} = (p_i)_{i \in [n]} \in [0, 1]^n$ , we define its *moment profile*  $m_{\mathcal{P}}$  to be the  $(2d(\epsilon) + 1)$ -dimensional vector

$$m_{\mathcal{P}} = \left( \sum_{i \in \mathcal{L}_{\mathcal{P}}} p_i, \sum_{i \in \mathcal{L}_{\mathcal{P}}} p_i^2, \dots, \sum_{i \in \mathcal{L}_{\mathcal{P}}} p_i^{d(\epsilon)}; \sum_{i \in \mathcal{R}_{\mathcal{P}}} p_i, \dots, \sum_{i \in \mathcal{R}_{\mathcal{P}}} p_i^{d(\epsilon)}; |\{i \mid p_i = 1\}| \right).$$

By the previous discussion, for two collections  $\mathcal{P}, \mathcal{Q}$ , if  $m_{\mathcal{P}} = m_{\mathcal{Q}}$  then  $d_{\text{TV}}(\text{PBD}(\mathcal{P}), \text{PBD}(\mathcal{Q})) \leq \epsilon$ .

Given the above we sparsify  $\mathcal{S}'_{n,\epsilon}$  as follows: for every possible moment profile that can arise from a Poisson Binomial distribution in  $k$ -sparse form, we keep in our cover a single Poisson Binomial distribution with such moment profile. The cover resulting from this sparsification is a  $2\epsilon$ -cover, since the sparsification loses us an additional  $\epsilon$  in total variation distance, as argued above.

We now bound the cardinality of the sparsified cover. The total number of moment profiles of  $k$ -sparse Poisson Binomial distributions is  $k^{O(d(\epsilon)^2)} \cdot (n + 1)$ . Indeed, consider a Poisson Binomial distribution  $\text{PBD}(\mathcal{P} = (p_i)_{i \in [n]})$  in  $k$ -sparse form. There are at most  $k^3 + 1$  choices for  $|\mathcal{L}_{\mathcal{P}}|$ , at

<sup>1</sup>Note that imposing the condition  $p_1 \leq \dots \leq p_\ell$  won't lose us any Poisson Binomial distribution in  $k$ -sparse form given Lemma 1.

most  $k^3 + 1$  choices for  $|\mathcal{R}_{\mathcal{P}}|$ , and at most  $(n + 1)$  choices for  $|\{i \mid p_i = 1\}|$ . We also claim that the total number of possible vectors

$$\left( \sum_{i \in \mathcal{L}_{\mathcal{P}}} p_i, \sum_{i \in \mathcal{L}_{\mathcal{P}}} p_i^2, \dots, \sum_{i \in \mathcal{L}_{\mathcal{P}}} p_i^{d(\epsilon)} \right)$$

is  $k^{O(d(\epsilon)^2)}$ . Indeed, if  $|\mathcal{L}_{\mathcal{P}}| = 0$  there is just one such vector, namely the all-zero vector. If  $|\mathcal{L}_{\mathcal{P}}| > 0$ , then, for all  $t = 1, \dots, d(\epsilon)$ ,  $\sum_{i \in \mathcal{L}_{\mathcal{P}}} p_i^t \in (0, |\mathcal{L}_{\mathcal{P}}|]$  and it must be an integer multiple of  $1/k^{2t}$ . So the total number of possible values of  $\sum_{i \in \mathcal{L}_{\mathcal{P}}} p_i^t$  is at most  $k^{2t} |\mathcal{L}_{\mathcal{P}}| \leq k^{2t} k^3$ , and the total number of possible vectors

$$\left( \sum_{i \in \mathcal{L}_{\mathcal{P}}} p_i, \sum_{i \in \mathcal{L}_{\mathcal{P}}} p_i^2, \dots, \sum_{i \in \mathcal{L}_{\mathcal{P}}} p_i^{d(\epsilon)} \right)$$

is at most

$$\prod_{t=1}^{d(\epsilon)} k^{2t} k^3 \leq k^{O(d(\epsilon)^2)}.$$

The same upper bound applies to the total number of possible vectors

$$\left( \sum_{i \in \mathcal{R}_{\mathcal{P}}} p_i, \sum_{i \in \mathcal{R}_{\mathcal{P}}} p_i^2, \dots, \sum_{i \in \mathcal{R}_{\mathcal{P}}} p_i^{d(\epsilon)} \right).$$

The moment profiles we enumerated over are a superset of the moment profiles of  $k$ -sparse Poisson Binomial distributions. We call them *compatible* moment profiles. We argued that there are at most  $k^{O(d(\epsilon)^2)} \cdot (n + 1)$  compatible moment profiles, so the total number of Poisson Binomial distributions in  $k$ -sparse form that we keep in the cover is at most  $k^{O(d(\epsilon)^2)} \cdot (n + 1) = n \cdot \left(\frac{1}{\epsilon}\right)^{O(\log^2 1/\epsilon)}$ . The number of Poisson Binomial distributions in  $(n, k)$ -Binomial form is the same as before, i.e. at most  $n^2$ , as we did not eliminate any of them. So the size of the sparsified cover is  $n^2 + n \cdot \left(\frac{1}{\epsilon}\right)^{O(\log^2 1/\epsilon)}$ .

To finish the proof it remains to argue that we don't actually need to first compute  $\mathcal{S}'_{n, \epsilon}$  and then sparsify it to obtain our cover, but can produce it directly in time  $O(n^2 \log n) + O(n \log n) \cdot \left(\frac{1}{\epsilon}\right)^{O(\log^2 1/\epsilon)}$ . We claim that, given a moment profile  $m$  that is compatible with a  $k$ -sparse Poisson Binomial distribution, we can compute some PBD  $(\mathcal{P} = (p_i)_i)$  in  $k$ -sparse form such that  $m_{\mathcal{P}} = m$ , if such a distribution exists, in time  $O(\log n) \left(\frac{1}{\epsilon}\right)^{O(\log^2 1/\epsilon)}$ . This follows from Claim 1 of Section 6.<sup>2</sup> So our algorithm enumerates over all moment profiles that are compatible with a  $k$ -sparse Poisson Binomial distribution and for each profile invokes Claim 1 to find a Poisson Binomial distribution with such moment profile, if such distribution exists, adding it to the cover if it does exist. It then enumerates over all Poisson Binomial distributions in  $(n, k)$ -Binomial form and adds them to the cover as well. The overall running time is as promised.

---

<sup>2</sup>A naive application of Claim 1 results in running time  $O(n^3 \log n) \cdot \left(\frac{1}{\epsilon}\right)^{O(\log^2 1/\epsilon)}$ . We can improve this to the claimed running time as follows: for all possible values  $|\mathcal{L}_{\mathcal{P}}|, |\mathcal{R}_{\mathcal{P}}|$  such that  $|\mathcal{L}_{\mathcal{P}}| + |\mathcal{R}_{\mathcal{P}}| \leq \min(k^3, n - m_{2d(\epsilon)+1})$ , we invoke Claim 1 with  $\tilde{n} = |\mathcal{L}_{\mathcal{P}}| + |\mathcal{R}_{\mathcal{P}}|$ ,  $\delta = d(\epsilon)$ ,  $B = k^3$ ,  $n_0 = n_1 = 0$ ,  $n_s = |\mathcal{L}_{\mathcal{P}}|$ ,  $n_b = |\mathcal{R}_{\mathcal{P}}|$ , and moments  $\mu_\ell = m_\ell$ , for  $\ell = 1, \dots, d(\epsilon)$ , and  $\mu'_\ell = m_{d(\epsilon)+\ell}$ , for  $\ell = 1, \dots, d(\epsilon)$ . If for some pair  $|\mathcal{L}_{\mathcal{P}}|, |\mathcal{R}_{\mathcal{P}}|$  the algorithm succeeds in finding probabilities matching the provided moments, we set  $m_{2d(\epsilon)+1}$  of the remaining probabilities equal to 1 and the rest to 0. Otherwise, we output "fail."

## 4 Proof of Theorem 2

We organize the proof according to the structure and notation of our outline in Section 1.2. In particular, we proceed to provide the details of Stages 1 and 2, described in the outline. The reader should refer to Section 1.2 for notation.

### 4.1 Details of Stage 1

Define  $\mathcal{L}_k := \{i \mid i \in [n] \wedge p_i \in (0, 1/k)\}$  and  $\mathcal{H}_k := \{i \mid i \in [n] \wedge p_i \in (1 - 1/k, 1)\}$ . We define the expectations  $(p'_i)_i$  of the intermediate indicators  $(Z_i)_i$  as follows.

First, we set  $p'_i = p_i$ , for all  $i \in [n] \setminus \mathcal{L}_k \cup \mathcal{H}_k$ . It follows that

$$d_{\text{TV}} \left( \sum_{i \in [n] \setminus \mathcal{L}_k \cup \mathcal{H}_k} X_i, \sum_{i \in [n] \setminus \mathcal{L}_k \cup \mathcal{H}_k} Z_i \right) = 0. \quad (5)$$

Next, we define the probabilities  $p'_i$ ,  $i \in \mathcal{L}_k$ , using the following procedure:

1. Set  $r = \left\lfloor \frac{\sum_{i \in \mathcal{L}_k} p_i}{1/k} \right\rfloor$ ; and let  $\mathcal{L}'_k \subseteq \mathcal{L}_k$  be an arbitrary subset of cardinality  $|\mathcal{L}'_k| = r$ .
2. Set  $p'_i = \frac{1}{k}$ , for all  $i \in \mathcal{L}'_k$ , and  $p'_i = 0$ , for all  $i \in \mathcal{L}_k \setminus \mathcal{L}'_k$ .

We bound the total variation distance  $d_{\text{TV}} \left( \sum_{i \in \mathcal{L}_k} X_i, \sum_{i \in \mathcal{L}_k} Z_i \right)$  using the Poisson approximation to the Poisson Binomial distribution. In particular, Theorem 4 implies

$$d_{\text{TV}} \left( \sum_{i \in \mathcal{L}_k} X_i, \text{Poisson} \left( \sum_{i \in \mathcal{L}_k} p_i \right) \right) \leq \frac{\sum_{i \in \mathcal{L}_k} p_i^2}{\sum_{i \in \mathcal{L}_k} p_i} \leq \frac{\frac{1}{k} \sum_{i \in \mathcal{L}_k} p_i}{\sum_{i \in \mathcal{L}_k} p_i} = 1/k.$$

Similarly,  $d_{\text{TV}} \left( \sum_{i \in \mathcal{L}_k} Z_i, \text{Poisson} \left( \sum_{i \in \mathcal{L}_k} p'_i \right) \right) \leq 1/k$ . Finally, we use Lemma 3 (given below and proved in Section 6) to bound the distance

$$d_{\text{TV}} \left( \text{Poisson} \left( \sum_{i \in \mathcal{L}_k} p_i \right), \text{Poisson} \left( \sum_{i \in \mathcal{L}_k} p'_i \right) \right) \leq \frac{1}{2} \left( e^{\frac{1}{k}} - e^{-\frac{1}{k}} \right) \leq \frac{1.5}{k},$$

where we used that  $|\sum_{i \in \mathcal{L}_k} p_i - \sum_{i \in \mathcal{L}_k} p'_i| \leq 1/k$ . Using the triangle inequality the above imply

$$d_{\text{TV}} \left( \sum_{i \in \mathcal{L}_k} X_i, \sum_{i \in \mathcal{L}_k} Z_i \right) \leq \frac{3.5}{k}. \quad (6)$$

**Lemma 3** (Variation Distance of Poisson Distributions). *Let  $\lambda_1, \lambda_2 > 0$ . Then*

$$d_{\text{TV}} (\text{Poisson}(\lambda_1), \text{Poisson}(\lambda_2)) \leq \frac{1}{2} \left( e^{|\lambda_1 - \lambda_2|} - e^{-|\lambda_1 - \lambda_2|} \right).$$

We follow a similar rounding scheme to define  $(p'_i)_{i \in \mathcal{H}_k}$  from  $(p_i)_{i \in \mathcal{H}_k}$ . That is, we round some of the  $p_i$ 's to  $1 - 1/k$  and some of them to 1 so that  $|\sum_{i \in \mathcal{H}_k} p_i - \sum_{i \in \mathcal{H}_k} p'_i| \leq 1/k$ . As a result, we get (to see this, repeat the argument employed above to the variables  $1 - X_i$  and  $1 - Z_i$ ,  $i \in \mathcal{H}_k$ )

$$d_{\text{TV}} \left( \sum_{i \in \mathcal{H}_k} X_i, \sum_{i \in \mathcal{H}_k} Z_i \right) \leq \frac{3.5}{k}. \quad (7)$$

Using (5), (6), (7) and Lemma 2 we get (2).

## 4.2 Details of Stage 2

Recall that  $\mathcal{M} := \{i \mid p'_i \notin \{0, 1\}\}$  and  $m := |\mathcal{M}|$ . Depending on whether  $m \leq k^3$  or  $m > k^3$  we follow different strategies to define the expectations  $(q_i)_i$  of indicators  $(Y_i)_i$ .

### 4.2.1 The Case $m \leq k^3$

First we set  $q_i = p'_i$ , for all  $i \in [n] \setminus \mathcal{M}$ . It follows that

$$d_{\text{TV}} \left( \sum_{i \in [n] \setminus \mathcal{M}} Z_i, \sum_{i \in [n] \setminus \mathcal{M}} Y_i \right) = 0. \quad (8)$$

For the definition of  $(q_i)_{i \in \mathcal{M}}$ , we make use of Ehm's Binomial approximation to the Poisson Binomial distribution, stated as Theorem 5 in Section 2.1. We start by partitioning  $\mathcal{M}$  as  $\mathcal{M} = \mathcal{M}_l \sqcup \mathcal{M}_h$ , where  $\mathcal{M}_l = \{i \in \mathcal{M} \mid p'_i \leq 1/2\}$ , and describe below a procedure for defining  $(q_i)_{i \in \mathcal{M}_l}$  so that the following hold:

1.  $d_{\text{TV}} \left( \sum_{i \in \mathcal{M}_l} Z_i, \sum_{i \in \mathcal{M}_l} Y_i \right) \leq 17/k$ ;
2. for all  $i \in \mathcal{M}_l$ ,  $q_i$  is an integer multiple of  $1/k^2$ .

To define  $(q_i)_{i \in \mathcal{M}_h}$ , we apply the same procedure to  $(1 - p'_i)_{i \in \mathcal{M}_h}$  to obtain  $(1 - q_i)_{i \in \mathcal{M}_h}$ . Assuming the correctness of our procedure for probabilities  $\leq 1/2$  the following should also hold:

1.  $d_{\text{TV}} \left( \sum_{i \in \mathcal{M}_h} Z_i, \sum_{i \in \mathcal{M}_h} Y_i \right) \leq 17/k$ ;
2. for all  $i \in \mathcal{M}_h$ ,  $q_i$  is an integer multiple of  $1/k^2$ .

Using Lemma 2, the above bounds imply

$$d_{\text{TV}} \left( \sum_{i \in \mathcal{M}} Z_i, \sum_{i \in \mathcal{M}} Y_i \right) \leq d_{\text{TV}} \left( \sum_{i \in \mathcal{M}_l} Z_i, \sum_{i \in \mathcal{M}_l} Y_i \right) + d_{\text{TV}} \left( \sum_{i \in \mathcal{M}_h} Z_i, \sum_{i \in \mathcal{M}_h} Y_i \right) \leq 34/k. \quad (9)$$

Now that we have (9), using (8) and Lemma 2 we get (3).

So it suffices to define the  $(q_i)_{i \in \mathcal{M}_l}$  properly. To do this, we define the partition  $\mathcal{M}_l = \mathcal{M}_{l,1} \sqcup \mathcal{M}_{l,2} \sqcup \dots \sqcup \mathcal{M}_{l,k-1}$  where for all  $j$ :

$$\mathcal{M}_{l,j} = \left\{ i \mid p'_i \in \left[ \frac{1}{k} + \frac{(j-1)j}{2} \frac{1}{k^2}, \frac{1}{k} + \frac{(j+1)j}{2} \frac{1}{k^2} \right) \right\}.$$

(Notice that the length of interval used in the definition of  $\mathcal{M}_{l,j}$  is  $\frac{j}{k^2}$ .) Now, for each  $j = 1, \dots, k-1$  such that  $\mathcal{M}_{l,j} \neq \emptyset$ , we define  $(q_i)_{i \in \mathcal{M}_{l,j}}$  via the following procedure:

1. Set  $p_{j,\min} := \frac{1}{k} + \frac{(j-1)j}{2} \frac{1}{k^2}$ ,  $p_{j,\max} := \frac{1}{k} + \frac{(j+1)j}{2} \frac{1}{k^2}$ ,  $n_j = |\mathcal{M}_{l,j}|$ ,  $\bar{p}_j = \frac{\sum_{i \in \mathcal{M}_{l,j}} p'_i}{n_j}$ .
2. Set  $r = \lfloor \frac{n_j(\bar{p}_j - p_{j,\min})}{j/k^2} \rfloor$ ; let  $\mathcal{M}'_{l,j} \subseteq \mathcal{M}_{l,j}$  be an arbitrary subset of cardinality  $r$ .
3. Set  $q_i = p_{j,\max}$ , for all  $i \in \mathcal{M}'_{l,j}$ ;
4. for an arbitrary index  $i_j^* \in \mathcal{M}_{l,j} \setminus \mathcal{M}'_{l,j}$ , set  $q_{i_j^*} = n_j \bar{p}_j - (rp_{j,\max} + (n_j - r - 1)p_{j,\min})$ ;
5. finally, set  $q_i = p_{j,\min}$ , for all  $i \in \mathcal{M}_{l,j} \setminus \mathcal{M}'_{l,j} \setminus \{i_j^*\}$ .

It is easy to see that

1.  $\sum_{i \in \mathcal{M}_{l,j}} p'_i = \sum_{i \in \mathcal{M}_{l,j}} q_i \equiv n_j \bar{p}_j$ ;
2. for all  $i \in \mathcal{M}_{l,j} \setminus \{i_j^*\}$ ,  $q_i$  is an integer multiple of  $1/k^2$ .

Moreover Theorem 5 implies:

$$d_{\text{TV}} \left( \sum_{i \in \mathcal{M}_{l,j}} Z_i, \mathcal{B}(n_j, \bar{p}_j) \right) \leq \frac{\sum_{i \in \mathcal{M}_{l,j}} (p'_i - \bar{p}_j)^2}{(n_j + 1)\bar{p}_j(1 - \bar{p}_j)} \leq \begin{cases} \frac{n_j(j \frac{1}{k^2})^2}{(n_j + 1)p_{j,\min}(1 - p_{j,\min})}, & \text{when } j < k - 1 \\ \frac{n_j(j \frac{1}{k^2})^2}{(n_j + 1)p_{j,\max}(1 - p_{j,\max})}, & \text{when } j = k - 1 \end{cases} \\ \leq \frac{8}{k^2}.$$

A similar derivation gives  $d_{\text{TV}} \left( \sum_{i \in \mathcal{M}_{l,j}} Y_i, \mathcal{B}(n_j, \bar{p}_j) \right) \leq \frac{8}{k^2}$ . So by the triangle inequality:

$$d_{\text{TV}} \left( \sum_{i \in \mathcal{M}_{l,j}} Z_i, \sum_{i \in \mathcal{M}_{l,j}} Y_i \right) \leq \frac{16}{k^2}. \quad (10)$$

As Eq (10) holds for all  $j = 1, \dots, k - 1$ , an application of Lemma 2 gives:

$$d_{\text{TV}} \left( \sum_{i \in \mathcal{M}_l} Z_i, \sum_{i \in \mathcal{M}_l} Y_i \right) \leq \sum_{j=1}^{k-1} d_{\text{TV}} \left( \sum_{i \in \mathcal{M}_{l,j}} Z_i, \sum_{i \in \mathcal{M}_{l,j}} Y_i \right) \leq \frac{16}{k}.$$

Moreover, the  $q_i$ 's defined above are integer multiples of  $1/k^2$ , except maybe for  $q_{i_1^*}, \dots, q_{i_{k-1}^*}$ . But we can round these to their closest multiple of  $1/k^2$ , increasing  $d_{\text{TV}} \left( \sum_{i \in \mathcal{M}_l} Z_i, \sum_{i \in \mathcal{M}_l} Y_i \right)$  by at most  $1/k$ .

#### 4.2.2 The Case $m > k^3$

Let  $t = |\{i \mid p'_i = 1\}|$ . We show that the random variable  $\sum_i Z_i$  is within total variation distance  $9/k$  from the Binomial distribution  $\mathcal{B}(m', q)$  where

$$m' := \left\lceil \frac{(\sum_{i \in \mathcal{M}} p'_i + t)^2}{\sum_{i \in \mathcal{M}} p_i'^2 + t} \right\rceil \quad \text{and} \quad q := \frac{\ell^*}{n},$$

where  $\ell^*$  satisfies  $\frac{\sum_{i \in \mathcal{M}} p'_i + t}{m'} \in [\frac{\ell^* - 1}{n}, \frac{\ell^*}{n}]$ . Notice that:

- $(\sum_{i \in \mathcal{M}} p'_i + t)^2 \leq (\sum_{i \in \mathcal{M}} p_i'^2 + t)(m + t)$ , by the Cauchy-Schwarz inequality; and
- $\frac{\sum_{i \in \mathcal{M}} p'_i + t}{m'} \leq \frac{\sum_{i \in \mathcal{M}} p'_i + t}{(\sum_{i \in \mathcal{M}} p_i'^2 + t)^2} = \frac{\sum_{i \in \mathcal{M}} p_i'^2 + t}{\sum_{i \in \mathcal{M}} p_i'^2 + t} \leq 1$ .

So  $m' \leq m + t \leq n$ , and there exists some  $\ell^* \in \{1, \dots, n\}$  so that  $\frac{\sum_{i \in \mathcal{M}} p_i'^2 + t}{m'} \in [\frac{\ell^* - 1}{n}, \frac{\ell^*}{n}]$ .

For fixed  $m'$  and  $q$ , we set  $q_i = q$ , for all  $i \leq m'$ , and  $q_i = 0$ , for all  $i > m'$ , and compare the distributions of  $\sum_{i \in \mathcal{M}} Z_i$  and  $\sum_{i \in \mathcal{M}} Y_i$ . For convenience we define

$$\begin{aligned} \mu &:= \mathbb{E} \left[ \sum_{i \in \mathcal{M}} Z_i \right] \quad \text{and} \quad \mu' := \mathbb{E} \left[ \sum_{i \in \mathcal{M}} Y_i \right], \\ \sigma^2 &:= \text{Var} \left[ \sum_{i \in \mathcal{M}} Z_i \right] \quad \text{and} \quad \sigma'^2 := \text{Var} \left[ \sum_{i \in \mathcal{M}} Y_i \right]. \end{aligned}$$

The following lemma compares the values  $\mu, \mu', \sigma, \sigma'$ .

**Lemma 4.** *The following hold*

$$\mu \leq \mu' \leq \mu + 1, \tag{11}$$

$$\sigma^2 - 1 \leq \sigma'^2 \leq \sigma^2 + 2, \tag{12}$$

$$\mu \geq k^2, \tag{13}$$

$$\sigma^2 \geq k^2 \left( 1 - \frac{1}{k} \right). \tag{14}$$

The proof of Lemma 4 is given in Section 6. To compare  $\sum_{i \in \mathcal{M}} Z_i$  and  $\sum_{i \in \mathcal{M}} Y_i$  we approximate both by Translated Poisson distributions. Theorem 6 implies that

$$\begin{aligned} d_{\text{TV}} \left( \sum_i Z_i, TP(\mu, \sigma^2) \right) &\leq \frac{\sqrt{\sum_i p_i'^3 (1 - p'_i)} + 2}{\sum_i p'_i (1 - p'_i)} \leq \frac{\sqrt{\sum_i p'_i (1 - p'_i)} + 2}{\sum_i p'_i (1 - p'_i)} \\ &\leq \frac{1}{\sqrt{\sum_i p'_i (1 - p'_i)}} + \frac{2}{\sum_i p'_i (1 - p'_i)} = \frac{1}{\sigma} + \frac{2}{\sigma^2} \\ &\leq \frac{1}{k \sqrt{1 - 1/k}} + \frac{2}{k^2 (1 - \frac{1}{k})} \quad (\text{using (14)}) \\ &\leq \frac{3}{k}, \end{aligned}$$

where for the last inequality we assumed  $k \geq 3$ , but the bound of  $3/k$  clearly also holds for  $k = 1, 2$ . Similarly,

$$\begin{aligned} d_{\text{TV}} \left( \sum_i Y_i, TP(\mu', \sigma'^2) \right) &\leq \frac{1}{\sigma'} + \frac{2}{\sigma'^2} \leq \frac{1}{k \sqrt{1 - \frac{1}{k} - \frac{1}{k^2}}} + \frac{2}{k^2 (1 - \frac{1}{k} - \frac{1}{k^2})} \quad (\text{using (12), (14)}) \\ &\leq \frac{3}{k}, \end{aligned}$$

where for the last inequality we assumed  $k \geq 3$ , but the bound of  $3/k$  clearly also holds for  $k = 1, 2$ . By the triangle inequality we then have that

$$\begin{aligned} d_{\text{TV}} \left( \sum_i Z_i, \sum_i Y_i \right) &\leq d_{\text{TV}} \left( \sum_i Z_i, TP(\mu, \sigma^2) \right) + d_{\text{TV}} \left( \sum_i Y_i, TP(\mu', \sigma'^2) \right) + d_{\text{TV}} (TP(\mu, \sigma^2), TP(\mu', \sigma'^2)) \\ &= 6/k + d_{\text{TV}} (TP(\mu, \sigma^2), TP(\mu', \sigma'^2)). \end{aligned} \quad (15)$$

It remains to bound the total variation distance between the two Translated Poisson distributions. We make use of the following lemma.

**Lemma 5** ([BL06]). *Let  $\mu_1, \mu_2 \in \mathbb{R}$  and  $\sigma_1^2, \sigma_2^2 \in \mathbb{R}_+ \setminus \{0\}$  be such that  $\lfloor \mu_1 - \sigma_1^2 \rfloor \leq \lfloor \mu_2 - \sigma_2^2 \rfloor$ . Then*

$$d_{\text{TV}} (TP(\mu_1, \sigma_1^2), TP(\mu_2, \sigma_2^2)) \leq \frac{|\mu_1 - \mu_2|}{\sigma_1} + \frac{|\sigma_1^2 - \sigma_2^2| + 1}{\sigma_1^2}.$$

Lemma 5 implies

$$\begin{aligned} d_{\text{TV}} (TP(\mu, \sigma^2), TP(\mu', \sigma'^2)) &\leq \frac{|\mu - \mu'|}{\min(\sigma, \sigma')} + \frac{|\sigma^2 - \sigma'^2| + 1}{\min(\sigma^2, \sigma'^2)} \\ &\leq \frac{1}{k \sqrt{1 - \frac{1}{k} - \frac{1}{k^2}}} + \frac{3}{k^2 (1 - \frac{1}{k} - \frac{1}{k^2})} \quad (\text{using Lemma 4}) \\ &\leq 3/k, \end{aligned} \quad (16)$$

where for the last inequality we assumed  $k > 3$ , but the bound clearly also holds for  $k = 1, 2, 3$ . Using (15) and (16) we get

$$d_{\text{TV}} \left( \sum_i Z_i, \sum_i Y_i \right) \leq 9/k, \quad (17)$$

which implies (3).

## 5 Proof of Theorem 3

Let  $\mathcal{X}$  and  $\mathcal{Y}$  be two collections of indicators as in the statement of Theorem 3. For  $\alpha_\ell(\cdot, \cdot)$  defined as in the statement of Theorem 7, we claim the following.

**Lemma 6.** *If  $\mathcal{P}, \mathcal{Q} \in [0, 1]^n$  satisfy property  $(C_d)$  in the statement of Theorem 3, then for all  $p, \ell \in \{0, \dots, d\}$ :*

$$\alpha_\ell(\mathcal{P}, p) = \alpha_\ell(\mathcal{Q}, p).$$

**Proof of lemma 6:** First  $\alpha_0(\mathcal{P}, p) = 1 = \alpha_0(\mathcal{Q}, p)$  by definition. Now fix  $\ell \in \{1, \dots, d\}$  and consider the function  $f(\vec{x}) := \alpha_\ell((x_1, \dots, x_n), p)$  in the variables  $x_1, \dots, x_n \in \mathbb{R}$ . Observe that  $f$  is a symmetric polynomial of degree  $\ell$  on  $x_1, \dots, x_n$ . Hence, from the theory of symmetric

polynomials, it follows that  $f$  can be written as a polynomial function of the power-sum symmetric polynomials  $\pi_1, \dots, \pi_\ell$ , where

$$\pi_j(x_1, \dots, x_n) := \sum_{i=1}^n x_i^j, \text{ for all } j \in [\ell],$$

as the elementary symmetric polynomial of degree  $j \in [n]$  can be written as a polynomial function of the power-sum symmetric polynomials  $\pi_1, \dots, \pi_j$  (e.g. [Zol87]). Now  $(C_d)$  implies that  $\pi_j(\mathcal{P}) = \pi_j(\mathcal{Q})$ , for all  $j \leq \ell$ . So  $f(\mathcal{P}) = f(\mathcal{Q})$ , i.e.  $\alpha_\ell(\mathcal{P}, p) = \alpha_\ell(\mathcal{Q}, p)$ . ■

For all  $p \in [0, 1]$ , by combining Theorem 7 and Lemma 6 and we get that

$$Pr[X = m] - Pr[Y = m] = \sum_{\ell=d+1}^n (\alpha_\ell(\mathcal{P}, p) - \alpha_\ell(\mathcal{Q}, p)) \cdot \delta^\ell \mathcal{B}_{n,p}(m), \text{ for all } m \in \{0, \dots, n\}.$$

Hence, for all  $p$ :

$$\begin{aligned} d_{\text{TV}}(X, Y) &= \frac{1}{2} \sum_{m=0}^n |Pr[X = m] - Pr[Y = m]| \\ &\leq \frac{1}{2} \sum_{\ell=d+1}^n |\alpha_\ell(\mathcal{P}, p) - \alpha_\ell(\mathcal{Q}, p)| \cdot \|\delta^\ell \mathcal{B}_{n,p}(\cdot)\|_1 \\ &\leq \frac{1}{2} \sum_{\ell=d+1}^n (|\alpha_\ell(\mathcal{P}, p)| + |\alpha_\ell(\mathcal{Q}, p)|) \cdot \|\delta^\ell \mathcal{B}_{n,p}(\cdot)\|_1. \end{aligned} \quad (18)$$

Plugging  $p = \bar{p} := \frac{1}{n} \sum_i p_i$  into Proposition 1, we get

$$\theta(\mathcal{P}, \bar{p}) = \frac{\sum_{i=1}^n (p_i - \bar{p})^2}{n\bar{p}(1 - \bar{p})} \leq \left| \max_i \{p_i\} - \min_i \{p_i\} \right| \leq \frac{1}{2} \quad (\text{see [Roo00]})$$

and then

$$\begin{aligned} \frac{1}{2} \sum_{\ell=d+1}^n |\alpha_\ell(\mathcal{P}, \bar{p})| \cdot \|\delta^\ell \mathcal{B}_{n,\bar{p}}(\cdot)\|_1 &\leq \sqrt{e}(d+1)^{1/4} 2^{-(d+1)/2} \frac{1 - \frac{1}{\sqrt{2}} \frac{d}{d+1}}{(\sqrt{2} - 1)^2} \\ &\leq 6.5(d+1)^{1/4} 2^{-(d+1)/2}. \end{aligned}$$

But  $(C_d)$  implies that  $\sum_i q_i = \sum_i p_i = \bar{p}$ . So we get in a similar fashion

$$\frac{1}{2} \sum_{\ell=d+1}^n |\alpha_\ell(\mathcal{Q}, \bar{p})| \cdot \|\delta^\ell \mathcal{B}_{n,\bar{p}}(\cdot)\|_1 \leq 6.5(d+1)^{1/4} 2^{-(d+1)/2}.$$

Plugging these bounds into (18) we get

$$d_{\text{TV}}(X, Y) \leq 13(d+1)^{1/4} 2^{-(d+1)/2}.$$

## 6 Deferred Proofs

**Proof of Lemma 1:** Let  $X = \sum_i X_i$  and  $Y = \sum_i Y_i$ . It is obvious that, if  $(p_1, \dots, p_n) = (q_1, \dots, q_n)$ , then the distributions of  $X$  and  $Y$  are the same. In the other direction, we show that, if  $X$  and  $Y$  have the same distribution, then  $(p_1, \dots, p_n) = (q_1, \dots, q_n)$ . Consider the polynomials:

$$g_X(s) = \mathbb{E}[(1+s)^X] = \prod_{i=1}^n \mathbb{E}[(1+s)^{X_i}] = \prod_{i=1}^n (1+p_i s);$$

$$g_Y(s) = \mathbb{E}[(1+s)^Y] = \prod_{i=1}^n \mathbb{E}[(1+s)^{Y_i}] = \prod_{i=1}^n (1+q_i s).$$

Since  $X$  and  $Y$  have the same distribution,  $g_X$  and  $g_Y$  are equal, so they have the same degree and roots. Notice that  $g_X$  has degree  $n - |\{i \mid p_i = 0\}|$  and roots  $\{-\frac{1}{p_i} \mid p_i \neq 0\}$ . Similarly,  $g_Y$  has degree  $n - |\{i \mid q_i = 0\}|$  and roots  $\{-\frac{1}{q_i} \mid q_i \neq 0\}$ . Hence,  $(p_1, \dots, p_n) = (q_1, \dots, q_n)$ . ■

**Proof of Lemma 2:** It follows from the coupling lemma that for any coupling of the variables  $X_1, \dots, X_n, Y_1, \dots, Y_n$ :

$$d_{\text{TV}} \left( \sum_{i=1}^n X_i, \sum_{i=1}^n Y_i \right) \leq \Pr \left[ \sum_{i=1}^n X_i \neq \sum_{i=1}^n Y_i \right]$$

$$\leq \sum_{i=1}^n \Pr[X_i \neq Y_i]. \quad (19)$$

We proceed to fix a specific coupling. For all  $i$ , it follows from the optimal coupling theorem that there exists a coupling of  $X_i$  and  $Y_i$  such that  $\Pr[X_i \neq Y_i] = d_{\text{TV}}(X_i, Y_i)$ . Using these individual couplings for each  $i$  we define a grand coupling of the variables  $X_1, \dots, X_n, Y_1, \dots, Y_n$  such that  $\Pr[X_i \neq Y_i] = d_{\text{TV}}(X_i, Y_i)$ , for all  $i$ . This coupling is faithful because  $X_1, \dots, X_n$  are mutually independent and  $Y_1, \dots, Y_n$  are mutually independent. Under this coupling Eq (19) implies:

$$d_{\text{TV}} \left( \sum_{i=1}^n X_i, \sum_{i=1}^n Y_i \right) \leq \sum_{i=1}^n \Pr[X_i \neq Y_i] \equiv \sum_{i=1}^n d_{\text{TV}}(X_i, Y_i). \quad (20)$$

■

**Claim 1.** Fix integers  $\tilde{n}, \delta, B, k \in \mathbb{N}_+$ ,  $\tilde{n}, k \geq 2$ . Given a set of values  $\mu_1, \dots, \mu_\delta, \mu'_1, \dots, \mu'_\delta$ , where, for all  $\ell = 1, \dots, \delta$ ,

$$\mu_\ell, \mu'_\ell \in \left\{ 0, \left(\frac{1}{k^2}\right)^\ell, 2\left(\frac{1}{k^2}\right)^\ell, \dots, B \right\},$$

discrete sets  $\mathcal{T}_1, \dots, \mathcal{T}_{\tilde{n}} \subseteq \{0, \frac{1}{k^2}, \frac{2}{k^2}, \dots, 1\}$ , and four integers  $n_0, n_1 \leq \tilde{n}$ ,  $n_s, n_b \leq B$ , it is possible to solve the system of equations:

$$(\Sigma) : \quad \sum_{p_i \in (0, 1/2]} p_i^\ell = \mu_\ell, \text{ for all } \ell = 1, \dots, \delta,$$

$$\sum_{p_i \in (1/2, 1)} p_i^\ell = \mu'_\ell, \text{ for all } \ell = 1, \dots, \delta,$$

$$|\{i \mid p_i = 0\}| = n_0$$

$$|\{i \mid p_i = 1\}| = n_1$$

$$|\{i \mid p_i \in (0, 1/2]\}| = n_s$$

$$|\{i \mid p_i \in (1/2, 1)\}| = n_b$$

with respect to the variables  $p_1 \in \mathcal{T}_1, \dots, p_{\tilde{n}} \in \mathcal{T}_{\tilde{n}}$ , or to determine that no solution exists, in time

$$O(\tilde{n}^3 \log_2 \tilde{n})B^{O(\delta)}k^{O(\delta^2)}.$$

**Proof of Claim 1:** We use dynamic programming. Let us consider the following tensor of dimension  $2\delta + 5$ :

$$A(i, z_0, z_1, z_s, z_b; \nu_1, \dots, \nu_\delta; \nu'_1, \dots, \nu'_\delta),$$

where  $i \in [\tilde{n}]$ ,  $z_0, z_1 \in \{0, \dots, \tilde{n}\}$ ,  $z_s, z_b \in \{0, \dots, B\}$  and

$$\nu_\ell, \nu'_\ell \in \left\{ 0, \left(\frac{1}{k^2}\right)^\ell, 2\left(\frac{1}{k^2}\right)^\ell, \dots, B \right\},$$

for  $\ell = 1, \dots, \delta$ . The total number of cells in  $A$  is

$$\tilde{n} \cdot (\tilde{n} + 1)^2 \cdot (B + 1)^2 \cdot \left( \prod_{\ell=1}^{\delta} (Bk^{2\ell} + 1) \right)^2 \leq O(\tilde{n}^3)B^{O(\delta)}k^{O(\delta^2)}.$$

Every cell of  $A$  is assigned value 0 or 1, as follows:

$$A(i, z_0, z_1, z_s, z_b; \nu_1, \dots, \nu_\delta, \nu'_1, \dots, \nu'_\delta) = 1$$

$$\iff \left( \begin{array}{l} \text{There exist } p_1 \in \mathcal{T}_1, \dots, p_i \in \mathcal{T}_i \text{ such} \\ \text{that } |\{j \leq i | p_j = 0\}| = z_0, \\ |\{j \leq i | p_j = 1\}| = z_1, \\ |\{j \leq i | p_j \in (0, 1/2]\}| = z_s, \\ |\{j \leq i | p_j \in (1/2, 1)\}| = z_b, \\ \sum_{j \leq i: p_j \in (0, 1/2]} p_j^\ell = \nu_\ell, \text{ for all} \\ \ell = 1, \dots, \delta, \sum_{j \leq i: p_j \in (1/2, 1)} p_j^\ell = \nu'_\ell, \text{ for} \\ \text{all } \ell = 1, \dots, \delta. \end{array} \right).$$

Notice that we need  $O(\tilde{n}^3)B^{O(\delta)}k^{O(\delta^2)}$  bits to store  $A$  and  $O(\log \tilde{n} + \delta \log B + \delta^2 \log k)$  bits to address cells of  $A$ . To complete  $A$  we can work in layers of increasing  $i$ . We initialize all entries to value 0. Then, the first layer  $A(1, \cdot, \cdot; \cdot, \dots, \cdot)$  can be completed easily as follows:

$$A(1, 1, 0, 0, 0; 0, 0, \dots, 0; 0, 0, \dots, 0) = 1 \iff 0 \in \mathcal{T}_1;$$

$$A(1, 0, 1, 0, 0; 0, 0, \dots, 0; 0, 0, \dots, 0) = 1 \iff 1 \in \mathcal{T}_1;$$

$$A(1, 0, 0, 1, 0; p, p^2, \dots, p^\delta; 0, \dots, 0) = 1 \iff p \in \mathcal{T}_1 \cap (0, 1/2];$$

$$A(1, 0, 0, 0, 1; 0, \dots, 0; p, p^2, \dots, p^\delta) = 1 \iff p \in \mathcal{T}_1 \cap (1/2, 1).$$

Inductively, to complete layer  $i + 1$ , we consider all the non-zero entries of layer  $i$  and for every such non-zero entry and for every  $v_{i+1} \in \mathcal{T}_{i+1}$ , we find which entry of layer  $i + 1$  we would transition to if we chose  $p_{i+1} = v_{i+1}$ . We set that entry equal to 1 and we also save a pointer to this entry from the corresponding entry of layer  $i$ , labeling that pointer with the value  $v_{i+1}$ . The bit operations required to complete layer  $i + 1$  are bounded by

$$|\mathcal{T}_{i+1}|(\tilde{n} + 1)^2 B^{O(\delta)} k^{O(\delta^2)} O(\log \tilde{n} + \delta \log B + \delta^2 \log k) \leq O(\tilde{n}^2 \log \tilde{n}) B^{O(\delta)} k^{O(\delta^2)}.$$

Therefore, the overall time needed to complete  $A$  is

$$O(\tilde{n}^3 \log \tilde{n}) B^{O(\delta)} k^{O(\delta^2)}.$$

Having completed  $A$ , it is easy to check if there is a solution to  $(\Sigma)$ . A solution exists if and only if

$$A(\tilde{n}, n_0, n_1, n_s, n_b; \mu_1, \dots, \mu_\delta; \mu'_1, \dots, \mu'_\delta) = 1,$$

and can be found by tracing the pointers from this cell of  $A$  back to level 1. The overall running time is dominated by the time needed to complete  $A$ . ■

**Proof of lemma 3:** Without loss of generality assume that  $0 < \lambda_1 \leq \lambda_2$  and denote  $\delta = \lambda_2 - \lambda_1$ . For all  $i \in \{0, 1, \dots\}$ , denote

$$p_i = e^{-\lambda_1} \frac{\lambda_1^i}{i!} \quad \text{and} \quad q_i = e^{-\lambda_2} \frac{\lambda_2^i}{i!}.$$

Finally, define  $\mathcal{I}^* = \{i : p_i \geq q_i\}$ .

We have

$$\begin{aligned} \sum_{i \in \mathcal{I}^*} |p_i - q_i| &= \sum_{i \in \mathcal{I}^*} (p_i - q_i) \leq \sum_{i \in \mathcal{I}^*} \frac{1}{i!} (e^{-\lambda_1} \lambda_1^i - e^{-\lambda_1 - \delta} \lambda_1^i) \\ &= \sum_{i \in \mathcal{I}^*} \frac{1}{i!} e^{-\lambda_1} \lambda_1^i (1 - e^{-\delta}) \\ &\leq (1 - e^{-\delta}) \sum_{i=0}^{+\infty} \frac{1}{i!} e^{-\lambda_1} \lambda_1^i = 1 - e^{-\delta}. \end{aligned}$$

On the other hand

$$\begin{aligned} \sum_{i \notin \mathcal{I}^*} |p_i - q_i| &= \sum_{i \notin \mathcal{I}^*} (q_i - p_i) \leq \sum_{i \notin \mathcal{I}^*} \frac{1}{i!} (e^{-\lambda_1} (\lambda_1 + \delta)^i - e^{-\lambda_1} \lambda_1^i) \\ &= \sum_{i \notin \mathcal{I}^*} \frac{1}{i!} e^{-\lambda_1} ((\lambda_1 + \delta)^i - \lambda_1^i) \\ &\leq \sum_{i=0}^{+\infty} \frac{1}{i!} e^{-\lambda_1} ((\lambda_1 + \delta)^i - \lambda_1^i) \\ &= e^\delta \sum_{i=0}^{+\infty} \frac{1}{i!} e^{-(\lambda_1 + \delta)} (\lambda_1 + \delta)^i - \sum_{i=0}^{+\infty} \frac{1}{i!} e^{-\lambda_1} \lambda_1^i \\ &= e^\delta - 1. \end{aligned}$$

Combining the above we get the result. ■

**Proof of lemma 4:** We have

$$\frac{\mu}{m'} = \frac{\sum_{i \in \mathcal{M}} p'_i + t}{m'} \leq q = \frac{\ell^*}{n} \leq \frac{\sum_{i \in \mathcal{M}} p'_i + t}{m'} + \frac{1}{n} = \frac{\mu}{m'} + \frac{1}{n}.$$

Multiplying by  $m'$  we get:

$$\mu \leq m'q \leq \mu + \frac{m'}{n}.$$

As  $\mu' = m'q$  and  $m' \leq n$ , we get  $\mu \leq \mu' \leq \mu + 1$ . Moreover, since  $m \geq k^3$ ,

$$\mu \geq \sum_{i \in \mathcal{M}} p'_i \geq m \frac{1}{k} \geq k^2.$$

For the variances we have:

$$\begin{aligned}
\sigma'^2 &= m'q(1-q) = m' \cdot \frac{\ell^*}{n} \cdot \left(1 - \frac{\ell^* - 1}{n} - \frac{1}{n}\right) \\
&\geq \left(\sum_{i \in \mathcal{M}} p'_i + t\right) \cdot \left(1 - \frac{1}{n} - \frac{\sum_{i \in \mathcal{M}} p'_i + t}{m'}\right) \\
&= (1 - 1/n) \left(\sum_{i \in \mathcal{M}} p'_i + t\right) - \frac{(\sum_{i \in \mathcal{M}} p'_i + t)^2}{m'} \\
&\geq (1 - 1/n) \left(\sum_{i \in \mathcal{M}} p'_i + t\right) - \frac{(\sum_{i \in \mathcal{M}} p'_i + t)^2}{\frac{(\sum_{i \in \mathcal{M}} p'_i + t)^2}{\sum_{i \in \mathcal{M}} p_i'^2 + t}} \\
&= \sum_{i \in \mathcal{M}} p'_i(1 - p'_i) - \frac{1}{n} \left(\sum_{i \in \mathcal{M}} p'_i + t\right) \\
&= \sigma^2 - \frac{1}{n} \left(\sum_{i \in \mathcal{M}} p'_i + t\right) \geq \sigma^2 - 1.
\end{aligned} \tag{21}$$

In the other direction:

$$\begin{aligned}
\sigma'^2 &= m'q(1-q) = m' \cdot \left(\frac{\ell^* - 1}{n} + \frac{1}{n}\right) \cdot \left(1 - \frac{\ell^*}{n}\right) \\
&\leq m' \cdot \left(\frac{\ell^* - 1}{n}\right) \cdot \left(1 - \frac{\ell^*}{n}\right) + \frac{m'}{n} \\
&\leq \left(\sum_{i \in \mathcal{M}} p'_i + t\right) \cdot \left(1 - \frac{\sum_{i \in \mathcal{M}} p'_i + t}{m'}\right) + 1 \\
&= \left(\sum_{i \in \mathcal{M}} p'_i + t\right) - \frac{(\sum_{i \in \mathcal{M}} p'_i + t)^2}{m'} + 1 \\
&\leq \left(\sum_{i \in \mathcal{M}} p'_i + t\right) - \frac{(\sum_{i \in \mathcal{M}} p'_i + t)^2}{\frac{(\sum_{i \in \mathcal{M}} p'_i + t)^2}{\sum_{i \in \mathcal{M}} p_i'^2 + t} + 1} + 1 \\
&= \left(\sum_{i \in \mathcal{M}} p'_i + t\right) - \left(\sum_{i \in \mathcal{M}} p_i'^2 + t\right) \frac{(\sum_{i \in \mathcal{M}} p'_i + t)^2}{(\sum_{i \in \mathcal{M}} p'_i + t)^2 + \sum_{i \in \mathcal{M}} p_i'^2 + t} + 1 \\
&= \left(\sum_{i \in \mathcal{M}} p'_i + t\right) - \left(\sum_{i \in \mathcal{M}} p_i'^2 + t\right) \left(1 - \frac{\sum_{i \in \mathcal{M}} p_i'^2 + t}{(\sum_{i \in \mathcal{M}} p'_i + t)^2 + \sum_{i \in \mathcal{M}} p_i'^2 + t}\right) + 1 \\
&= \sum_{i \in \mathcal{M}} p'_i(1 - p'_i) + \frac{(\sum_{i \in \mathcal{M}} p_i'^2 + t)^2}{(\sum_{i \in \mathcal{M}} p'_i + t)^2 + \sum_{i \in \mathcal{M}} p_i'^2 + t} + 1 \\
&= \sigma^2 + \frac{(\sum_{i \in \mathcal{M}} p_i'^2 + t)^2}{(\sum_{i \in \mathcal{M}} p'_i + t)^2 + \sum_{i \in \mathcal{M}} p_i'^2 + t} + 1 \leq \sigma^2 + 2.
\end{aligned} \tag{22}$$

Finally,

$$\sigma^2 = \sum_{i \in \mathcal{M}} p'_i(1 - p'_i) \geq m \frac{1}{k} \left(1 - \frac{1}{k}\right) \geq k^2 \left(1 - \frac{1}{k}\right).$$

■

**Proposition 2.** For all  $d \in [n]$ , Condition  $(C_d)$  in the statement of Theorem 3 is equivalent to the following condition:

$$(V_d) : \mathbb{E} \left[ \left( \sum_{i=1}^n X_i \right)^\ell \right] = \mathbb{E} \left[ \left( \sum_{i=1}^n Y_i \right)^\ell \right], \text{ for all } \ell \in [d].$$

**Proof of Proposition 2:**

$(V_d) \Rightarrow (C_d)$ : First notice that, for all  $\ell \in [n]$ ,  $\mathbb{E} \left[ \left( \sum_{i=1}^n X_i \right)^\ell \right]$  can be written as a weighted sum of the elementary symmetric polynomials  $\psi_1(\mathcal{P}), \psi_2(\mathcal{P}), \dots, \psi_\ell(\mathcal{P})$ , where, for all  $t \in [n]$ ,  $\psi_t(\mathcal{P})$  is defined as

$$\psi_t(\mathcal{P}) := (-1)^t \sum_{\substack{S \subseteq [n] \\ |S| = t}} \prod_{i \in S} p_i.$$

$(V_d)$  implies then by induction

$$\psi_\ell(\mathcal{P}) = \psi_\ell(\mathcal{Q}), \quad \text{for all } \ell = 1, \dots, d. \quad (23)$$

Next, for all  $t \in [n]$ , define  $\pi_t(\mathcal{P})$  to be the power sum symmetric polynomial of degree  $t$

$$\pi_t(\mathcal{P}) := \sum_{i=1}^n p_i^t.$$

Now, fix any  $\ell \leq d$ . Since  $\pi_\ell(\mathcal{P})$  is a symmetric polynomial of degree  $\ell$  on the variables  $p_1, \dots, p_n$ , it can be expressed as a function of the elementary symmetric polynomials  $\psi_1(\mathcal{P}), \dots, \psi_\ell(\mathcal{P})$ . So, by (23),  $\pi_\ell(\mathcal{P}) = \pi_\ell(\mathcal{Q})$ . Since this holds for any  $\ell \leq d$ ,  $(C_d)$  is satisfied.

The implication  $(C_d) \Rightarrow (V_d)$  is established in a similar fashion.  $(C_d)$  says that

$$\pi_\ell(\mathcal{P}) = \pi_\ell(\mathcal{Q}), \text{ for all } \ell = 1, \dots, d. \quad (24)$$

Fix some  $\ell \leq d$ .  $\mathbb{E} \left[ \left( \sum_{i=1}^n X_i \right)^\ell \right]$  can be written as a weighted sum of the elementary symmetric polynomials  $\psi_1(\mathcal{P}), \psi_2(\mathcal{P}), \dots, \psi_\ell(\mathcal{P})$ . Also, for all  $t \in [\ell]$ ,  $\psi_t(\mathcal{P})$  can be written as a polynomial function of  $\pi_1(\mathcal{P}), \dots, \pi_t(\mathcal{P})$  (see, e.g., [Zol87]). So from (24) it follows that  $\mathbb{E} \left[ \left( \sum_{i=1}^n X_i \right)^\ell \right] = \mathbb{E} \left[ \left( \sum_{i=1}^n Y_i \right)^\ell \right]$ . Since this holds for any  $\ell \leq d$ ,  $(V_d)$  is satisfied. ■

**Corollary 1.** Let  $\mathcal{P} := (p_i)_{i=1}^n \in [1/2, 1]^n$  and  $\mathcal{Q} := (q_i)_{i=1}^n \in [1/2, 1]^n$  be two collections of probability values in  $[1/2, 1]$ . Let also  $\mathcal{X} := (X_i)_{i=1}^n$  and  $\mathcal{Y} := (Y_i)_{i=1}^n$  be two collections of mutually independent indicators with  $\mathbb{E}[X_i] = p_i$  and  $\mathbb{E}[Y_i] = q_i$ , for all  $i \in [n]$ . If for some  $d \in [n]$  Condition  $(C_d)$  in the statement of Theorem 3 is satisfied, then

$$d_{\text{TV}} \left( \sum_i X_i, \sum_i Y_i \right) \leq 13(d+1)^{1/4} 2^{-(d+1)/2}.$$

**Proof of Corollary 1:** Define  $X'_i = 1 - X_i$  and  $Y'_i = 1 - Y_i$ , for all  $i$ . Also, denote  $p'_i = \mathbb{E}[X'_i] = 1 - p_i$  and  $q'_i = \mathbb{E}[Y'_i] = 1 - q_i$ , for all  $i$ . By assumption:

$$\sum_{i=1}^n (1 - p'_i)^\ell = \sum_{i=1}^n (1 - q'_i)^\ell, \quad \text{for all } \ell = 1, \dots, d. \quad (25)$$

Using the Binomial theorem and induction, we see that (25) implies:

$$\sum_{i=1}^n p_i^\ell = \sum_{i=1}^n q_i^\ell, \quad \text{for all } \ell = 1, \dots, d.$$

Hence we can apply Theorem 3 to deduce

$$d_{\text{TV}} \left( \sum_i X'_i, \sum_i Y'_i \right) \leq 13(d+1)^{1/4} 2^{-(d+1)/2}.$$

The proof is completed by noticing that

$$d_{\text{TV}} \left( \sum_i X_i, \sum_i Y_i \right) = d_{\text{TV}} \left( \sum_i X'_i, \sum_i Y'_i \right).$$

■

## Acknowledgement

We thank the anonymous reviewer for comments that helped improve the presentation.

## References

- [Ber41] Andrew C. Berry. The Accuracy of the Gaussian Approximation to the Sum of Independent Variates. *Transactions of the American Mathematical Society*, 49(1):122–136, 1941. [1.4](#)
- [BH84] Andrew D. Barbour and Peter Hall. On the Rate of Poisson Convergence. *Mathematical Proceedings of the Cambridge Philosophical Society*, 95(03):473–480, 1984. [1.4](#), [4](#)
- [BJH92] Andrew D. Barbour, Lars Holst, and Svante Janson. *Poisson Approximation*. Oxford University Press, 1992. [4](#)
- [BL06] Andrew D. Barbour and Torgny Lindvall. Translated Poisson Approximation for Markov Chains. *Journal of Theoretical Probability*, 19(3):609–630, 2006. [5](#)
- [Blo99] Matthias Blonski. Anonymous Games with Binary Actions. *Games and Economic Behavior*, 28(2):171–180, 1999. [1](#)
- [CGS10] Louis H. Y. Chen, Larry Goldstein, and Qi-Man Shao. *Normal Approximation by Stein's Method*. Springer, 2010. [1.4](#)
- [Che74] Louis H. Y. Chen. On the Convergence of Poisson Binomial to Poisson Distributions. *Annals of Probability*, 2:178–180, 1974. [1.4](#)

- [CL97] Song X. Chen and Jun S. Liu. Statistical Applications of the Poisson-Binomial and Conditional Bernoulli Distributions. *Statistica Sinica*, 7:875–892, 1997. [1.4](#)
- [Das08] Constantinos Daskalakis. An Efficient PTAS for Two-Strategy Anonymous Games. In *the 4th International Workshop on Internet and Network Economics (WINE)*, 2008. [1.4](#)
- [DDS12] Constantinos Daskalakis, Ilias Diakonikolas, and Rocco Servedio. Learning Poisson Binomial Distributions. In *the 44th Annual ACM Symposium on Theory of Computing (STOC)*, 2012. [1.4](#)
- [DP86] Paul Deheuvels and Dietmar Pfeifer. A Semigroup Approach to Poisson Approximation. *Annals of Probability*, 14:663–676, 1986. [1.4](#)
- [DP07] Constantinos Daskalakis and Christos H. Papadimitriou. Computing Equilibria in Anonymous Games. In *the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2007. [1](#)
- [DP09] Constantinos Daskalakis and Christos H. Papadimitriou. On Oblivious PTAS’s for Nash Equilibrium. In *the 41st Annual ACM Symposium on Theory of Computing (STOC)*, 2009. [1](#), [1.4](#)
- [DP13] Constantinos Daskalakis and Christos H. Papadimitriou. Approximate Nash Equilibria in Anonymous Games. *Journal of Economic Theory*, to appear, 2013. [1](#)
- [Ehm91] Werner Ehm. Binomial Approximation to the Poisson Binomial Distribution. *Statistics and Probability Letters*, 11:7–16, 1991. [1.2](#), [1.4](#), [5](#)
- [Ess42] Carl-Gustav Esseen. On the Liapunoff Limit of Error in the Theory of Probability. *Arkiv för Matematik, Astronomi och Fysik*, A28:1–19, 1942. [1.4](#)
- [HC60] Joseph L. Hodges and Lucien Le Cam. The Poisson Approximation to the Poisson Binomial Distribution. *The Annals of Mathematical Statistics*, 31(3):737–740, 1960. [1.4](#)
- [LC60] Lucien Le Cam. An Approximation Theorem for the Poisson Binomial Distribution. *Pacific Journal of Mathematics*, 10:1181–1197, 1960. [1.4](#)
- [Mik93] Vladimir G. Mikhailov. On a Refinement of the Central Limit Theorem for Sums of Independent Random Indicators. *Theory of Probability and its Applications*, 38:479–489, 1993. [1.4](#)
- [Mil96] Igal Milchtaich. Congestion Games with Player-Specific Payoff Functions. *Games and Economic Behavior*, 13:111–124, 1996. [1](#)
- [MR95] Rajeev Motwani and Prabhakar Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995. [1.4](#)
- [Poi37] Siméon D. Poisson. *Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile*. Bachelier, Paris, 1837. [1.4](#)
- [Rö07] Adrian Röllin. Translated Poisson Approximation Using Exchangeable Pair Couplings. *Annals of Applied Probability*, 17:1596–1614, 2007. [1.2](#), [6](#)

- [Roo00] Bero Roos. Binomial Approximation to the Poisson Binomial Distribution: The Krawtchouk Expansion. *Theory of Probability and its Applications*, 45(2):328–344, 2000. [1.3](#), [1.4](#), [2.1](#), [7](#), [2.1](#), [1](#), [5](#)
- [Soo96] S.Y.T. Soon. Binomial Approximation for Dependent Indicators. *Statistica Sinica*, 6:703–714, 1996. [1.4](#)
- [Ste94] J. Michael Steele. Le Cam’s Inequality and Poisson Approximation. *American Mathematical Monthly*, 101:48–54, 1994. [1.4](#)
- [Vol95] A. Yu. Volkova. A Refinement of the Central Limit Theorem for Sums of Independent Random Indicators. *Theory of Probability and its Applications*, 40:791–794, 1995. [1.4](#)
- [Wan93] Yuan H. Wang. On the Number of Successes in Independent Trials. *Statistica Sinica*, 3:295–312, 1993. [1.4](#)
- [Zol87] Vladimir M. Zolotarev. Random Symmetric Polynomials. *Journal of Mathematical Sciences*, 38(5):2262–2272, 1987. [5](#), [6](#)