Abstract

While some useful probability bounds for the sum of $n$ pairwise independent Bernoulli random variables exceeding an integer $k$ have been proposed in the literature, none of these bounds are tight in general. In this paper, we provide three results towards finding tight probability bounds for this class of problems. Firstly, when $k = 1$, the tightest upper bound on the probability of the union of $n$ pairwise independent events is provided in closed form for any input marginal probability vector $\mathbf{p} \in [0, 1]^n$. Building on this, we show that the ratio of Boole’s union bound and the tight pairwise independent bound is upper bounded by $4/3$ and this bound is attained. Secondly, for $n$ pairwise independent events with identical probabilities $p \in [0, 1]$, the tightest upper bound on the probability that at least $k$ occur is provided in closed form. Lastly, for non-identical pairwise independent events, new upper bounds are derived for any $k \geq 2$. While these bounds are not always tight, they improve on existing bounds and in special instances are shown to be tight. The proofs of tightness are developed using techniques of linear optimization and should be of independent interest. Numerical examples are provided to illustrate when the bounds provide significant improvement over existing bounds.

1 Introduction

Probability bounds for sums of Bernoulli random variables have been extensively studied by researchers in various communities including, but not limited to, probability and statistics, computer science, combinatorics and optimization. In this paper, our focus is on pairwise independent Bernoulli random variables. It is well known that while mutually independent random variables are pairwise independent, the reverse is not true. Feller [1968] attributes Bernstein [1946] with identifying one of the earliest examples with $n = 3$ random variables which are pairwise independent, but not mutually independent. For general $n$, constructions of pairwise independent Bernoulli random variables can be found in the works of Geisser and Mantel [1962], Karloff and Mansour [1994], Koller and Meggido [1994], pairwise

*Engineering Systems and Design, Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372. Email: arjun_ramachandra@mymail.sutd.edu.sg
†Engineering Systems and Design, Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372. Email: karthik_natarajan@sutd.edu.sg
‡The research of the second author was partly supported by the MOE Academic Research Fund Tier 2 grant MOE2019-T2-2-138, “Enhancing Robustness of Networks to Dependence via Optimization”.

1
independent discrete random variables in Feller [1959], Lancaster [1965], Joffe [1974], O’Brien [1980] and pairwise independent normal random variables in Geisser and Mantel [1962]. One of the motivations for studying constructions of pairwise independent random variables particularly in the computer science community is that the joint distribution can have a low cardinality support (polynomial in the number of random variables) in comparison to mutually independent random variables (exponential in the number of random variables). The reader is referred to Lancaster [1965] and more recent papers of Babai [2013] and Gavinsky and Pudlák [2016] who have developed lower bounds on the entropy of the joint distribution of pairwise independent random variables which are shown to grow logarithmically with the number of random variables. The low cardinality of these distributions have important ramifications in efficiently derandomizing algorithms for NP-hard combinatorial optimization problems (see the review article of Luby and Widgerson [2005] and the references therein for results on pairwise independent and more generally $t$-wise independent random variables).

In this paper, we are interested in the problem of computing tight probability bounds for the sum of pairwise independent Bernoulli random variables. Given an integer $n \geq 2$, denote by $[n] = \{1, 2, \ldots, n\}$, the set of indices and by $K_n = \{(i,j) : 1 \leq i < j \leq n\}$, the set of all pairwise indices in $[n]$ (it can be viewed as a complete graph on $n$ nodes). Given integers $i < j$, let $[i,j] = \{i, i+1, \ldots, j-1, j\}$. Consider a Bernoulli random vector $\tilde{c} = (\tilde{c}_1, \ldots, \tilde{c}_n)$ with marginal probabilities given by $p_i = \mathbb{P}(\tilde{c}_i = 1)$ for $i \in [n]$. Denote by $p = (p_1, \ldots, p_n) \in [0,1]^n$, the univariate marginal vector and by $\Theta(\{0,1\}^n)$, the set of all probability distributions supported on $\{0,1\}^n$. Consider the set of joint probability distributions of Bernoulli random variables consistent with the given marginal probabilities and pairwise independence:

$$
\Theta(p, p_i p_j; (i,j) \in K_n) = \left\{ \theta \in \Theta(\{0,1\}^n) \mid \mathbb{P}_\theta(\tilde{c}_i = 1) = p_i, \forall i \in [n], \quad \mathbb{P}_\theta(\tilde{c}_i = 1, \tilde{c}_j = 1) = p_i p_j, \forall (i,j) \in K_n \right\}.
$$

This set of distributions is clearly nonempty for any $p \in [0,1]^n$, since the mutually independent distribution lies in the set. Our problem of interest is to compute the maximum probability that the sum of $n$ random variables exceeds an integer $k \in [n]$ for distributions in this set. Denote the tightest upper bound by $P(n,k,p)$ (observe that the bivariate probabilities here are simply given by the product of the univariate probabilities). Then,

$$
P(n,k,p) = \max_{\theta \in \Theta(p, p_i p_j; (i,j) \in K_n)} \mathbb{P}_\theta \left( \sum_{i=1}^{n} \tilde{c}_i \geq k \right).
$$

Two useful upper bounds that have been proposed for this problem are the following:

(a) Chebyshev [1867] bound: The one-sided version of the Chebyshev tail probability bound for any random variable uses only the mean and variance of the random variable. Since the Bernoulli random variables are assumed to be pairwise independent or equivalently uncorrelated, the variance of the sum is given by:

$$
\text{Variance} \left( \sum_{i=1}^{n} \tilde{c}_i \right) = \sum_{i=1}^{n} p_i (1 - p_i).
$$
Applying the classical Chebyshev bound then gives an upper bound:

\[
P(n, k, p) \leq \begin{cases} 
1, & k < \sum_{i=1}^{n} p_i, \\
\frac{\sum_{i=1}^{n} p_i (1 - p_i)}{\left(\sum_{i=1}^{n} p_i (1 - p_i) + (k - \sum_{i=1}^{n} p_i)^2\right)}, & \sum_{i=1}^{n} p_i \leq k \leq n.
\end{cases} \tag{1.1}
\]

(b) Schmidt et al. [1995] bound: The Schmidt, Siegel and Srinivasan bound is derived by bounding the tail probability using the moments of multilinear polynomials. This is in contrast to the Chernoff-Hoeffding bound (see Chernoff [1952], Hoeffding [1963]) which bounds the tail probability of the sum of independent random variables using the moment generating function. A multilinear polynomial of degree \(j\) in \(n\) variables is defined as:

\[
S_j(c) = \sum_{1 \leq i_1 < i_2 < ... < i_j \leq n} c_{i_1} c_{i_2} ... c_{i_j}.
\]

At the crux of their analysis is the observation that all the higher moments of the sum of Bernoulli random variables can be generated from linear combinations of the expected values of multilinear polynomials of the random variables. The construction of the bound makes use of the equality:

\[
\left(\sum_{i \in [n]} c_i^j\right) = S_j(c), \quad \forall c \in \{0, 1\}^n, \forall j \in [0, n], \tag{1.2}
\]

where \(S_0(c) = 1\) and \(^r \choose s = r!/(s!(r-s)!\) for any pair of integers \(r \geq s \geq 0\). The bound derived in Schmidt et al. [1995] (see Theorem 7, part (II) on page 239) for pairwise independent random variables is given by:\(^1:\)

\[
\bar{P}(n, k, p) \leq \min\left(1, \frac{\sum_{i \in [n]} p_i}{k}, \frac{\sum_{(i,j) \in K_n} p_i p_j}{\binom{k}{2}}\right). \tag{1.3}
\]

While both the bounds in (1.1) and (1.3) have shown to be useful and are easy to use, neither of them are tight for general values of \(n, k\) and \(p \in [0, 1]^n\). In this paper, we work towards identifying instances for pairwise independent random variables where these bounds can be tightened.

### 1.1 Other related bounds

Consider the set of joint probability distributions of Bernoulli random variables consistent with the marginal probability vector \(p \in [0, 1]^n\) and general bivariate probabilities given by \(p_{ij} = P(\tilde{c}_i = 1, \tilde{c}_j = 1)\) for \((i, j) \in K_n:\)

\[
\Theta(p, p_{ij}; (i, j) \in K_n) = \left\{ \theta \in \Theta([0, 1]^n) \mid P_\theta (\tilde{c}_i = 1) = p_i, \forall i \in [n], \quad P_\theta (\tilde{c}_i = 1, \tilde{c}_j = 1) = p_{ij}, \forall (i, j) \in K_n \right\}.
\]

\(^1:\)While the statement in the theorem in Schmidt et al. [1995] is for \(k > \sum_i p_i\), it is straightforward to see that their analysis would lead to the form here for general \(k\).
Unlike the pairwise independent case, verifying if this set of distributions is nonempty is already known to be a NP-complete problem (see Pitowsky [1991]). The tightest upper bound on the probability for distributions in this set is given by \( \max_{\theta \in \Theta(p,p_{ij};(i,j) \in K_n)} \mathbb{P}_\theta \left( \sum_{i=1}^n \tilde{c}_i \geq k \right) \) where the bound is set to \(-\infty\) if the set of feasible distributions is empty. The bound is given by the optimal value of the linear program (see Hailperin [1965]):

\[
\begin{align*}
\max & \sum_{c \in \{0,1\}^n : \sum_t c_t \geq k} \theta(c) \\
\text{s.t} & \sum_{c \in \{0,1\}^n} \theta(c) = 1,
\sum_{c \in \{0,1\}^n : c_i = 1} \theta(c) = p_i, \quad \forall i \in [n],
\sum_{c \in \{0,1\}^n : c_i = 1, c_j = 1} \theta(c) = p_{ij}, \quad \forall (i,j) \in K_n,
\theta(c) \geq 0, \quad \forall c \in \{0,1\}^n,
\end{align*}
\]

where the decision variables are the joint probabilities \( \theta(c) = \mathbb{P}(\tilde{c} = c) \) for all \( c \in \{0,1\}^n \). The number of decision variables in this formulation, however, grows exponentially in the number of random variables \( n \). The dual linear program is given by:

\[
\begin{align*}
\min & \sum_{(i,j) \in K_n} \lambda_{ij}p_{ij} + \sum_{i=1}^n \lambda_ip_i + \lambda_0 \\
\text{s.t} & \sum_{(i,j) \in K_n} \lambda_{ij}c_ic_j + \sum_{i=1}^n \lambda_ic_i + \lambda_0 \geq 0, \quad \forall c \in \{0,1\}^n,
\sum_{(i,j) \in K_n} \lambda_{ij}c_ic_j + \sum_{i=1}^n \lambda_ic_i + \lambda_0 \geq 1, \quad \forall c \in \{0,1\}^n : \sum_t c_t \geq k.
\end{align*}
\]

The dual linear program in (1.5) has a polynomial number of decision variables, exponential number of constraints and is always feasible. Strong duality hence holds. Given the large size of the primal and dual linear programs, two main approaches to tackle these problems have been studied in the literature:

i) The first approach is to find closed-form bounds by generating dual feasible solutions (see Kounias [1968], Kounias and Marin [1976], Sathe et al. [1980], Móri and Székely [1985], Dawson and Sankoff [1967], Galambos [1975, 1977], de Caen [1997], Kuai et al. [2000], Dohmen and Tittmann [2007] and related graph-based bounds in Hunter [1976], Worsley [1982], Veneziani [2008b], Vizvári [2007]). These bounds have shown to be tight in special instances (see Section 2.1 for examples).

ii) The second approach is to try and reduce the size of the linear programs using relaxations and to solve it numerically. Since the primal linear program in (1.4) quickly becomes intractable with an increase in the number of random variables \( n \), many papers adopting this approach, aggregate the primal decision variables, thus obtaining weaker bounds as a trade-off for the reduced size.
Formulations of linear programs under assumptions of partially or fully aggregated univariate, bivariate or \(m\)-variate information for \(2 \leq m < n\) have been proposed in Kwerel [1975b], Platz [1985], Prékopa [1988, 1990], Boros and Prékopa [1989], Prékopa and Gao [2005], Qiu et al. [2016], Yang et al. [2016], Yoda and Prékopa [2016]). In some cases closed-form bounds have been additionally derived for these aggregated linear programs (see Boros and Prékopa [1989]). Techniques to solve the dual formulations by restricting the dual variables have been similarly been studied (see Boros et al. [2014]).

To the best of our knowledge, the connection of these bounds which assume general bivariate information with tight bounds for pairwise independent random variables have not been well-studied in the literature. Another upper bound derived under weaker assumptions is the Boole [1854] (see also Fréchet [1935]) union bound \((k = 1)\) which is valid with arbitrary dependence among the Bernoulli random variables. Boole’s union bound is given as:

\[
P_u(n, 1, p) = \max_{\theta \in \Theta(p)} \mathbb{P}_\theta \left( \sum_{i=1}^{n} \tilde{c}_i \geq 1 \right) = \min \left( \sum_{i=1}^{n} p_i, 1 \right),
\]

where \(\Theta(p)\) denotes the set of joint distributions supported on \(\{0, 1\}^n\) consistent with the given univariate information:

\[
\Theta(p) = \left\{ \theta \in \Theta(\{0, 1\}^n) \mid \mathbb{P}_\theta (\tilde{c}_i = 1) = p_i, \forall i \in [n] \right\}.
\]

Clearly, \(P(n, 1, p) \leq P_u(n, 1, p)\). Extensions of this for \(k \geq 2\) is provided in Rüger [1978].

### 1.2 Contributions and structure

This brings us to the key contributions and the structure of the current paper:

(a) In Section 2, we provide the tightest upper bound on the probability on the union of \(n\) pairwise independent events \(P(n, 1, p)\) in closed form (see Theorem 2.1). The contributions of Theorem 2.1 lie in:

i) Establishing that when the random variables are pairwise independent, for any given marginal vector \(p \in [0, 1]^n\), the upper bound proposed in Kounias [1968], Hunter [1976] and Worsley [1982] is tight using techniques from linear optimization. These bounds were initially developed for the sum of dependent Bernoulli random variables with arbitrary bivariate probabilities using tree structures from graph theory and are not in general guaranteed to be tight (see Example 1 in Section 2.1). Interestingly for pairwise independent random variables, we prove that this bound is indeed tight.

ii) Constructing an extremal distribution (see Tables 1 and 2) which attains the bound appears to be nontrivial. The proof of tightness requires us to show that a feasible correlated distribution of a Bernoulli random vector \(\tilde{c}\) with an arbitrary univariate probability vector \(p \in [0, 1]^n\) and
transformed bivariate probabilities $p_ip_j/p$ where $p \in [\max_i p_i, 1]$, always exists. This feasibility problem is of independent interest in itself, since feasibility is typically not guaranteed for arbitrary correlation structures with Bernoulli random vectors.

iii) Building on the result (see Proposition 2.1), we show that the ratio of Boole’s union bound and the pairwise independent bound is upper bounded by $4/3$ and this is attained. Applications of the result in correlation gap analysis are discussed.

(b) In Section 3, for any $k \in [n]$, we provide the tightest upper bound in closed form when the marginals of the pairwise independent Bernoulli random variables are identical (see Theorem 3.1). The proof is based on showing an equivalence with a linear programming formulation of an aggregated moment bound for which closed-form solutions have been derived by Boros and Prékopa [1989]. While the tight closed-form bound is more complicated than the closed-form Chebyshev bound in (1.1) and the Schmidt, Siegel and Srinivasan bound in (1.3), it helps us identify conditions under which these relatively simpler bounds are guaranteed to be tight (see Proposition 3.1).

(c) In Section 4, for $k \geq 2$ and general probabilities $p \in [0,1]^n$, we develop new bounds using the results from Section 3 (see Theorem 4.1). These bounds improve on existing closed-form bounds and are shown to be tight in the special case when $n-1$ marginal probabilities are identical (see Proposition 4.1). Numerical examples are provided to quantify the improvement of the bounds over existing bounds.

(d) We conclude in Section 5 and identify some future research questions.

2 Tight upper bound for $k = 1$

The first theorem provides the tightest upper bound on the probability of the union of pairwise independent events. Towards this, we start by applying the bound derived in Kounias [1968] for this problem. Sort the probabilities in increasing value as $0 \leq p_1 \leq p_2 \leq \ldots \leq p_n \leq 1$. Consider the following solution to the dual linear program in (1.5) where $k = 1$ and $p_{ij} = p_ip_j$: 

$$
\lambda_0 = 0, \quad \lambda_i = 1 \forall i \in [n], \quad \lambda_{in} = -1 \forall i \in [n-1] \text{ and } \lambda_{ij} = 0 \text{ otherwise.}
$$

The left hand side of the dual constraints in (1.5) simplifies to:

$$
\sum_{(i,j) \in K_n} \lambda_{ij} c_i c_j + \sum_{i=1}^{n} \lambda_i c_i + \lambda_0 = - \sum_{i=1}^{n-1} c_i c_n + \sum_{i=1}^{n} c_i = c_n + \sum_{i=1}^{n-1} c_i (1 - c_n).
$$

To verify that this solution is dual feasible, observe that with all $c_i = 0$, $c_n + \sum_{i=1}^{n-1} c_i (1 - c_n) = 0$. When $c_n = 1$, regardless of the values of $c_1, \ldots, c_{n-1}$, we have $c_n + \sum_{i=1}^{n-1} c_i (1 - c_n) = 1$. Lastly, when $c_n = 0$
Theorem 2.1. Sort the probabilities in increasing value as $c_i$ for $i \in [n-1]$, we have $c_n + \sum_{i=1}^{n-1} c_i (1 - c_n) \geq 1$. This gives a dual feasible solution with the objective value $\sum_{i=1}^{n} p_i - p_n \left( \sum_{i=1}^{n-1} p_i \right)$. Another dual feasible solution for the linear program is given by:

$$\lambda_0 = 1, \; \lambda_i = 0 \; \forall i \in [n], \; \lambda_{ij} = 0 \; \forall (i, j) \in K_n,$$

with a dual objective of 1. From weak duality, we then have:

$$\mathcal{P}(n, 1, p) \leq \min \left( \sum_{i=1}^{n} p_i - p_n \left( \sum_{i=1}^{n-1} p_i \right), 1 \right).$$

The key result we show next is that there is always a feasible distribution which attains this upper bound. A simple construction of the extremal distribution appears to be tricky. In the proof, we show that this problem can be transformed to showing the existence of a distribution of a Bernoulli random variable and simulating from these distributions can be shown to always exist. While there are several results on finding specific correlation structures which are compatible with given Bernoulli random variables and simulating from these distributions (see Chaganty and Joe [2006], Qaqish [2003], Emrich and Piedmonte [1991], Lunn and Davies [1998]), to the best of our knowledge this particular case does not appear to be covered in these results. This brings us to the first theorem.

**Theorem 2.1.** Sort the probabilities in increasing value as $0 \leq p_1 \leq p_2 \leq \ldots \leq p_n \leq 1$. Then,

$$\overline{\mathcal{P}}(n, 1, p) = \min \left( \sum_{i=1}^{n} p_i - p_n \left( \sum_{i=1}^{n-1} p_i \right), 1 \right).$$

**Proof.** When $p_{ij} = p_i p_j$ and $k = 1$, the optimal value of the primal linear program in (1.4) is clearly bounded since feasibility is guaranteed and the objective function is a probability value. The optimality conditions of linear programming states that $\{\theta(c); c \in \{0, 1\}^n\}$ is primal optimal and $\{\lambda_{ij}; (i, j) \in K_n, \lambda_i; i \in [n], \lambda_0\}$ is dual optimal if and only if they satisfy: (i) the primal feasibility conditions in (1.4), (ii) the dual feasibility conditions in (1.5) and (iii) the complementary slackness conditions given by:

$$\left( \sum_{(i,j) \in K_n} \lambda_{ij} c_i c_j + \sum_{i=1}^{n} \lambda_i c_i + \lambda_0 \right) \theta(c) = 0, \; \forall c \in \{0, 1\}^n : \sum_t c_t = 0,$$

$$\left( \sum_{(i,j) \in K_n} \lambda_{ij} c_i c_j + \sum_{i=1}^{n} \lambda_i c_i + \lambda_0 - 1 \right) \theta(c) = 0, \; \forall c \in \{0, 1\}^n : \sum_t c_t \geq 1.$$

1) **Proof of tightness of non-trivial bound in (2.1)**

We now show that $\overline{\mathcal{P}}(n, 1, p) = \sum_{i=1}^{n} p_i - p_n \left( \sum_{i=1}^{n-1} p_i \right)$ which is the non-trivial part of the upper bound in (2.1) when $\sum_{i=1}^{n-1} p_i \leq 1$.

Step (1a): Show tightness by constructing a pairwise independent distribution
We verify the tightness of the bound, by showing there exists a primal solution (feasible distribution) which satisfies the complementary slackness conditions. Towards this, observe that from the complementary slackness condition in (iii):

\[ \forall c \in \{0, 1\}^n : \sum_{t=1}^{n-1} c_t \geq 2, c_n = 0, \text{we have} \quad \left( c_n + \sum_{i=1}^{n-1} c_i (1 - c_n) - 1 \right) > 0 \implies \theta(c) = 0. \]

This forces a total of \(2^{n-1} - n\) scenarios to have zero probability. Building on this, we set the probabilities of the \(2^n\) possible scenarios of \(\tilde{c}\) as shown in Table 1. The probability of the vector of all zeros (one scenario) is set to \(1 - \sum_{i=1}^{n} p_i + p_n \left( \sum_{i=1}^{n-1} p_i \right)\). To match the bivariate probabilities \(P(\tilde{c}_i = 1, \tilde{c}_n = 0) = p_i (1 - p_n)\), we have to then set the probability of the scenario where \(c_i = 1, c_n = 0\) and all remaining \(c_j = 0\) to \(p_i (1 - p_n)\). This corresponds to the \(n - 1\) scenarios in Table 1. Hence, to ensure feasibility of the distribution, we need to show that there exist nonnegative values of \(\theta(c)\) for the last \(2^{n-1}\) scenarios such that:

\[
\sum_{c \in \{0, 1\}^n : c_n = 1} \theta(c) = p_n,
\]

\[
\sum_{c \in \{0, 1\}^n : c_i = 1, c_n = 1} \theta(c) = p_i p_n, \quad \forall i \in [n - 1],
\]

\[
\sum_{c \in \{0, 1\}^n : c_i = 1, c_j = 1} \theta(c) = p_i p_j, \quad \forall (i, j) \in K_{n-1}.
\]

or equivalently, by conditioning on \(c_n = 1\), we need to show that there exists nonnegative joint
probabilities $\theta_{n-1}(c)$ where $\theta_{n-1}(c) = \mathbb{P}(\tilde{c} = c)$ for all $c \in \{0,1\}^{n-1}$ such that:

\[
\begin{align*}
\sum_{c \in \{0,1\}^{n-1}} \theta_{n-1}(c) &= 1, \\
\sum_{c \in \{0,1\}^{n-1} : c_i = 1} \theta_{n-1}(c) &= p_i, \quad \forall i \in [n-1], \\
\sum_{c \in \{0,1\}^{n-1} : c_i = 1, c_j = 1} \theta_{n-1}(c) &= \frac{p_ip_j}{p_n}, \quad \forall (i,j) \in K_{n-1},
\end{align*}
\]

This corresponds to verifying the existence of a probability distribution on $n-1$ Bernoulli random variables with univariate probabilities $p_i$ and bivariate probabilities $p_ip_j/p_n$ where $p_n \geq p_{n-1} \geq p_{n-2} \geq \ldots \geq p_1$. Observe, that in (2.2), the univariate probabilities remain the same but the random variables are no longer pairwise independent. In the next step of the proof, we show that such a distribution always exists.

**Step (1b): Show there exists a distribution that satisfies (2.2)**

(i) We first argue that it is sufficient to verify the existence of nonnegative values $\theta_{n-1}(c)$ for $n-1$ Bernoulli random variables such that:

\[
\begin{align*}
\sum_{c \in \{0,1\}^{n-1}} \theta_{n-1}(c) &= 1, \\
\sum_{c \in \{0,1\}^{n-1} : c_i = 1} \theta_{n-1}(c) &= p_i, \quad \forall i \in [n-1], \\
\sum_{c \in \{0,1\}^{n-1} : c_i = 1, c_j = 1} \theta_{n-1}(c) &= \frac{p_ip_j}{p_n}, \quad \forall (i,j) \in K_{n-1},
\end{align*}
\]

where the bivariate probabilities are modified from $p_ip_j/p_n$ to $p_ip_j/p_{n-1}$. To see this, since $1 \leq 1/p_n \leq 1/p_{n-1}$, it is always possible to find a $\lambda \in [0, 1]$ such that:

\[
\frac{1}{p_n} = \lambda \frac{1}{p_{n-1}} + (1 - \lambda)(1).
\]

Then, by considering a convex combination of the following two distributions:

\[
\theta_{n-1} = \lambda \overline{\theta}_{n-1} + (1 - \lambda)\theta_{n-1},
\]

where $\overline{\theta}_{n-1}$ is a probability distribution which satisfies (2.3) and $\theta_{n-1}$ is a pairwise independent joint distribution on $n-1$ Bernoulli random variables with univariate probabilities given by $p_i$ and bivariate probabilities given by $p_ip_j$, we guarantee (2.2) is feasible. The distribution $\theta_{n-1}$ always exists (simply choose the mutually independent distribution on $n-1$ random variables with univariate probabilities $p_i$). The distribution $\theta_{n-1}$ thus created will have univariate probabilities given by $p_i$ and bivariate probabilities given by $p_ip_j/p_n$.

(ii) To show that (2.3) is feasible, by conditioning on $c_{n-1} = 1$, we need to show that there
exists a probability distribution on \( n - 2 \) Bernoulli random variables with nonnegative values
\[
\theta_{n-2}(c) = P(\hat{c} = c) \text{ for all } c \in \{0, 1\}^{n-2}
\]
such that:
\[
\sum_{c \in \{0, 1\}^{n-2}} \theta_{n-2}(c) = 1, \\
\sum_{c \in \{0, 1\}^{n-2} : c_i = 1} \theta_{n-2}(c) = \frac{p_i}{p_n-1}, \quad \forall i \in [n-2], \\
\sum_{c \in \{0, 1\}^{n-2} : c_i = 1, c_j = 1} \theta_{n-2}(c) = \frac{p_ip_j}{p_n-1}, \quad \forall (i, j) \in K_{n-2}. 
\] (2.4)

Then, by setting the vector of all zeros to \( 1 - p_{n-1} \) and scaling the probabilities when \( c_{n-1} = 1 \) (see the construction in Table 2), we obtain a feasible distribution. A pairwise independent joint distribution \( \theta_{n-2} \) on \( n - 2 \) random variables specified by (2.4) with univariate probabilities given by \( p_i/p_{n-1} \) and bivariate probabilities given by \( (p_i/p_{n-1})(p_j/p_{n-1}) \) always exists (simply choose the mutually independent distribution on \( n - 2 \) random variables with univariate probabilities \( p_i/p_{n-1} \)). An outline of the different distributions used in the construction in steps (2) and (3) are shown in Figure 1. This completes the proof for the case where \( \sum_{i=1}^{n-1} p_i \leq 1 \) and the tight bound is given by:
\[
\mathcal{P}(n, 1, p) = \sum_{i=1}^{n} p_i - p_n \left(\sum_{i=1}^{n-1} p_i\right).
\]

Table 2: Probabilities of the scenarios \( (\theta_{n-1}(c)) \) to create a feasible distribution in (2.3).

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>( c_1 )</th>
<th>( c_2 )</th>
<th>...</th>
<th>( c_{n-1} )</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>2^{n-2} scenarios</td>
<td>0 0 ... 0</td>
<td>( \theta_{n-1}(c) = 1 - p_{n-1} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 0 ... 0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 ... 1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2^{n-2} scenarios</td>
<td>0 0 ... 1</td>
<td>( \theta_{n-1}(c) = p_{n-1}\theta_{n-2}(c) )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 1 1</td>
<td>( \theta_{n-1}(c) = p_{n-1}\theta_{n-2}(c) )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2) Proof of the tightness of trivial part of the bound in (2.1)
To complete the proof, consider the case with $\sum_{i=1}^{n-1} p_i > 1$. Then, there exists an index $t \in [2, n-1]$ such that $\sum_{i=1}^{t-1} p_i \leq 1$ and $\sum_{i=1}^{t} p_i > 1$. Let $\delta = 1 - \sum_{i=1}^{t-1} p_i$. Clearly $0 \leq \delta < p_t$. From steps (1)-(2) in the proof of the non-trivial bound, we know that there exists a distribution for $t + 1$ pairwise independent random variables with marginal probabilities $p_1, p_2, \ldots, p_{t-1}, \delta, p_{t+1}$ such that the probability of the sum of the random variables being at least one is equal to one (since the sum of the first $t$ probabilities in this case is equal to one). By increasing the marginal probability $\delta$ to $p_t$, we can only increase this probability. Hence, there exists a distribution for $t + 1$ pairwise independent random variables with probabilities $0 \leq p_1 \leq p_2 \leq \ldots \leq p_t \leq p_{t+1} \leq 1$ such that there is a zero probability of these random variables to simultaneously take a value of 0. We can generate the remaining random variables $\tilde{c}_{t+2}, \ldots, \tilde{c}_n$ independently with marginal probabilities $p_{t+2}, \ldots, p_n$. This provides a feasible distribution that attains the bound of one, thus completing the proof.

2.1 Connection of Theorem 2.1 to existing results
The problem of bounding the probability that the sum of Bernoulli random variables is at least one has been extensively studied in the literature, under knowledge of general bivariate probabilities. Let $A_i$ denote the event that $\tilde{c}_i = 1$ for each $i$, then, $k = 1$ simply corresponds to bounding the probability of the union of events. When the marginal probabilities $p_i = \mathbb{P}(A_i)$ for $i \in [n]$ and bivariate probabilities $p_{ij} = \mathbb{P}(A_i \cap A_j)$ for $(i, j) \in K_n$ are given, Hunter [1976] and Worsley [1982] derived the following bound
by optimizing over the spanning trees $\tau \in T$:

$$\mathbb{P}(\bigcup_i A_i) \leq \sum_{i=1}^{n} p_i - \max_{\tau \in T} \sum_{(i,j) \in \tau} p_{ij},$$

(2.5)

where $T$ is the set of all spanning trees on the complete graph with $n$ nodes (where the edge weights are given by $p_{ij}$). A special case of the Hunter [1976] bound was derived by Kounias [1968] as:

$$\mathbb{P}(\bigcup_i A_i) \leq \sum_{i=1}^{n} p_i - \max_{j \in [n]} \sum_{i \neq j} p_{ij},$$

(2.6)

which subtracts the maximum weight of a star spanning tree on the complete graph from the sum of the marginal probabilities $\sum_i p_i$. Tree bounds have been shown to be tight, in some special cases as outlined below:

i) **Zero bivariate probabilities for all pairs** ($p_{ij} = 0$, $\forall (i,j) \in K_n$):
   - When all the probabilities $p_{ij}$ are zero, the bound reduces to Boole’s union bound which is tight.

ii) **Zero bivariate probabilities outside a given tree**:
   - Given a tree $\tau$ such that the bivariate probabilities $p_{ij}$ are zero if and only if the edge $(i,j) \notin \tau$, Worsley [1982] proved that the bound is tight (see Veneziani [2008a] for related results).

iii) **Lower bounds on bivariate probabilities**:
   - Boros et al. [2014] proved that by relaxing the equality of bivariate probabilities to lower bounds on bivariate probabilities as
     $$\mathbb{P}(A_i \cap A_j) \geq p_{ij}, \forall (i,j) \in K_n,$$
   the tightest upper bound on the probability of the union is exactly the Hunter [1976] and Worsley [1982] bound (see Maurer [1983] for related results).

iv) **Pairwise independent variables (Theorem 2.1 in this paper)**:
   - With pairwise independent random variables where $p_{ij} = p_i p_j$, the maximum weight spanning trees in (2.5) is exactly the star tree with the root at node $n$ and edges $(i,n)$ for all $i \in [n-1]$.
   - In this case, the Kounias [1968], Hunter [1976] and Worsley [1982] bound reduces to the bound in (2.1) which is shown to be tight in Theorem 2.1 in this paper.

The next example illustrates that with general bivariate probabilities, even if a joint distribution exists, the Hunter [1976] and Worsley [1982] bound and thus the Kounias [1968] bound is not guaranteed to be tight.

*Example 1.* Consider $n = 4$ Bernoulli random variables with univariate marginal vector

$$p = [0.35, 0.19, 0.13, 0.2],$$

which reduces to the Hunter [1976] and Worsley [1982] bound and thus the Kounias [1968] bound is not guaranteed to be tight.
and bivariate probabilities

\[ p_{12} = 0.001, \quad p_{13} = 0.022, \quad p_{14} = 0.03, \quad p_{23} = 0.017, \quad p_{24} = 0.018, \quad p_{34} = 0.019. \]

It can be verified that a joint distribution with these given univariate and bivariates exists. The tight upper bound on the probability by solving the linear program is given by

\[
\max_{\theta \in \Theta(p, p_{ij}; (i, j) \in K_4)} \mathbb{P}_\theta (\tilde{c}_1 + \tilde{c}_2 + \tilde{c}_3 + \tilde{c}_4 \geq 1) = 0.784.
\]

Figure 2 displays the star spanning tree chosen by the Kounias [1968] bound and spanning tree chosen by the Hunter [1976] and Worsley [1982] bound. It is clear that none of these bounds are tight in this given instance. Boros et al. [2014] also provide randomly generated instances (see Table 1 of Section 4 in their paper) when the Hunter [1976] bound is not tight though it is the best performing among the upper bounds considered there.

![Figure 2](image)

**Figure 2:** Kounias [1968], Hunter [1976] and Worsley [1982] spanning trees with general bivariates

Figure 3 demonstrates that with the same set of univariate marginals \( p = [0.35, 0.19, 0.13, 0.2] \), when pairwise independence is enforced, both the Kounias [1968] and Hunter [1976] and Worsley [1982] spanning trees are identical and the bounds in (2.6) and (2.5) equal the tight bound 0.688 (from Theorem 2.1).
2.2 Comparison with the union bound

The next proposition provide an upper bound on the ratio of Boole’s union bound and the pairwise independent bound in Theorem 2.1.

**Proposition 2.1.** For all \( p \in [0, 1]^n \), we have:

\[
\frac{\mathcal{P}_u(n, 1, p)}{\mathcal{P}(n, 1, p)} \leq \frac{4}{3}.
\]

The ratio of 4/3 is attained when \( \sum_{i=1}^{n-1} p_i = 1/2 \) and \( p_n = 1/2 \).

**Proof.** Assume the probabilities are sorted in increasing values as \( 0 \leq p_1 \leq p_2 \leq \ldots \leq p_n \leq 1 \). It is straightforward to see that if \( \sum_{i=1}^{n-1} p_i > 1 \), both the bounds take the value \( \mathcal{P}(n, 1, p) = \mathcal{P}_u(n, 1, p) = 1 \). Now assume, \( \alpha = \sum_{i=1}^{n-1} p_i \leq 1 \). The ratio is given as:

\[
\frac{\mathcal{P}_u(n, 1, p)}{\mathcal{P}(n, 1, p)} = \frac{\min (\sum_{i=1}^{n} p_i, 1)}{\sum_{i=1}^{n} p_i - p_n \left( \sum_{i=1}^{n-1} p_i \right)} = \frac{\min (\alpha + p_n, 1)}{\alpha + p_n - \alpha p_n}.
\]
If $\alpha + p_n \leq 1$, then we have:
\[
\frac{P_u(n, 1, p)}{P(n, 1, p)} = \frac{\alpha + p_n}{\alpha + p_n - \alpha p_n} = \frac{1}{1 - \frac{1}{\alpha + \frac{1}{p_n}}} \leq \frac{4}{3}
\]
[where the maximum value is attained at $\alpha = 1 - p_n$ and $p_n = 1/2$].

If $\alpha + p_n \geq 1$, then we have:
\[
\frac{P_u(n, 1, p)}{P(n, 1, p)} = \frac{1}{\alpha + p_n - \alpha p_n} = \frac{1}{\alpha(1 - p_n) + p_n} \leq \frac{4}{3}
\]
[where the maximum value is attained at $\alpha = 1 - p_n$ and $p_n = 1/2$].

This gives the bound of $4/3$ when $p_n = 1/2$ and $\alpha = 1/2$.

We now illustrate an application of Theorem 2.1 and Proposition 2.1 in comparing bounds with dependent and independent random variables in correlation gap analysis.

**Example 2 (Correlation gap analysis).** The notion of a “correlation gap” was introduced by Agrawal et al. [2012]. It is defined as the ratio of the worst-case expected cost for random variables with given univariate marginals to the expected cost when the random variables are independent. When $\tilde{c}$ is a Bernoulli random vector and $\theta_{\text{ind}}$ denotes the independent distribution, the correlation gap is defined there as:

$$
\kappa_u(p) = \sup_{\theta \in \Theta(p)} \frac{\mathbb{E}_\theta[f(\tilde{c})]}{\mathbb{E}_{\theta_{\text{ind}}}[f(\tilde{c})]}.
$$

A key result in this area is that for any nonnegative, nondecreasing, submodular set function, $f(S)$, the correlation gap is always upper bounded by $e/(e - 1)$ (see Calinescu et al. [2007], Agrawal et al. [2012]). The example constructed in these papers to show this bound is attained is for the maximum of binary variables:

$$
f(c) = \max \{c_i \mid i \in [n]\}.
$$

This defines a nonnegative, nondecreasing, submodular set function $f(S)$ which takes a value zero when
\[ S = \emptyset \text{ and one when } S \neq \emptyset. \] For a given marginal vector \( p \), the correlation gap in (2.7) reduces to

\[
\kappa_u(p) = \frac{\max_{\theta \in \Theta(p)} \mathbb{E}_\theta[\max(\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_n)]}{1 - \prod_{i=1}^n (1 - p_i)}
\]

\[
= \frac{\max_{\theta \in \Theta(p)} \mathbb{P}_\theta(\sum_{i=1}^n \tilde{c}_i \geq 1)}{1 - \prod_{i=1}^n (1 - p_i)}
\]

\[
= \min \left( \sum_{i=1}^n p_i, 1 \right) \frac{1 - \prod_{i=1}^n (1 - p_i)}{1 - \prod_{i=1}^n (1 - p_i)}.
\]

(2.8)

We now provide an extension of this definition by considering the ratio of the worst-case expected cost when the random variables are pairwise independent to the expected cost when the random variables are independent. This is given as:

\[
\kappa(p) = \sup_{\theta \in \Theta(p, p_{ij}; (i, j) \in K_n)} \frac{\mathbb{E}_\theta[f(c)]}{\mathbb{E}_{\theta_{\text{ind}}}[f(c)]},
\]

which reduces in this specific case to:

\[
\kappa(p) = \min \left( \sum_{i=1}^n p_i - p_n \left( \sum_{i=1}^{n-1} p_i \right), 1 \right) \frac{1 - \prod_{i=1}^n (1 - p_i)}{1 - \prod_{i=1}^n (1 - p_i)}.
\]

Clearly \( \kappa(p) \leq \kappa_u(p) \). We now discuss behavior of these two ratios.

i) **Worst case analysis:**

Assume the marginal probability vector is given by \( p = (1/n, 1/n, \ldots, 1/n) \). For the independent distribution, the probability is given by \( 1 - (1 - 1/n)^n \), while Boole’s union bound is equal to one (attained by the distribution which assigns probability \( 1/n \) to each of \( n \) support points with \( c_i = 1, c_j = 0, \forall j \neq i \) (for each \( i \in [n] \)) and zero otherwise). In this case the limit of the ratio as \( n \) goes to infinity is given by:

\[
\lim_{n \to \infty} \kappa_u(p) = \frac{1}{1 - (1 - 1/n)^n} \to \frac{e}{e - 1} \approx 1.5819.
\]

Likewise it is easy to verify that with pairwise independence:

\[
\lim_{n \to \infty} \kappa(p) = \frac{1 - \frac{1}{n} \left( 1 - \frac{1}{n} \right)}{1 - (1 - 1/n)^n} = \frac{e}{e - 1} \approx 1.5819.
\]

Thus in the worst-case, both these bounds attain the ratio \( e/(e - 1) \).

ii) **Instances where correlation gap can be improved:**

On the other hand, Proposition 2.1 illustrates that for the probabilities \( p_n = 1/2 \) and \( \sum_{i=1}^{n-1} p_i = 1/2 \), the pairwise independent bound is \( 3/4 \) and Boole’s union bound is one. For example with
$n = 2$ where $p = (1/2, 1/2)$, Boole’s union bound is one, while both the pairwise independent and the independent probabilities are equal to $3/4$. Then, we have $\kappa_u((1/2, 1/2)) = 4/3$ while $\kappa((1/2, 1/2)) = 1$. Thus in specific instances, the correlation gap can be tightened by considering pairwise independent random variables.

3 Bounds with identical marginals

In this section, we provide probability bounds for the sum of pairwise independent random variables exceeding $k \in [2, n]$ where the marginals are identical. We observe that both Chebyshev’s bound and the Schmidt, Siegel and Srinivasan bound introduced in Section 1 can be expressed in terms of the first two aggregated (or equivalently binomial) moments for the sum of pairwise independent random variables, $S_1(p) = \mathbb{E}[S_1(\tilde{c})] = \sum_i p_i$ and $S_2(p) = \mathbb{E}[S_2(\tilde{c})] = \sum_{(i,j) \in K_n} p_i p_j$. The use of aggregated moments in developing probability bounds for sums of Bernoulli random variables has been analyzed in a series of papers by Prékopa [1988, 1990], Boros and Prékopa [1989], Prékopa and Gao [2005], Boros et al. [2014]. These papers use polynomial sized linear optimization formulations to find probability bounds and in some cases, closed-form bounds are derived. Boros and Prékopa [1989] derived the tightest upper bound on $P(\tilde{\xi} \geq k)$ over all distributions $\omega$ of an integer random variable $\tilde{\xi}$ (supported on $[0, n]$), which are assumed to lie in a set of distributions given by:

$$\left\{ \omega([0, n]) \mid \mathbb{E}_\omega \left[ \binom{\tilde{\xi}}{j} \right] = S_j, \ j = 1, 2 \right\}.$$  

The upper bound is a closed-form expression as follows:

$$P\left( \sum_{i=1}^{n} \tilde{c}_i \geq k \right) \leq \begin{cases} 1, & k < \frac{(n-1)S_1 - 2S_2}{n - S_1} \\ \frac{(k + n - 1)S_1 - 2S_2}{kn}, & k \geq 1 + \frac{2S_2}{S_1}, i = \left\lceil \frac{(k - i)S_1 - 2S_2}{k - S_1} \right\rceil, \\ \frac{(i - 1)(i - 2S_1) + 2S_2}{(k - i)^2 + (k - i)}, & k \geq 1 + \frac{2S_2}{S_1}, i = \left\lceil \frac{(k - 1)S_1 - 2S_2}{k - S_1} \right\rceil, \\ \frac{(n-1)S_1 - 2S_2}{n - S_1} \leq k < 1 + \frac{2S_2}{S_1} & \end{cases} \quad (3.1)$$

where the ceiling function $\lceil x \rceil$ maps $x$ to the smallest integer greater than or equal to $x$. The bound in (3.1) can be applied to pairwise independent variables with $\tilde{\xi} = \sum_{i \in [n]} \tilde{c}_i$. The next theorem provides the tight bound with identical marginals, the proof of which is based on showing equivalence of the exponential sized linear program which computes the exact bound with the polynomial sized linear program analyzed in computing the bound in (3.1).
Theorem 3.1. Assume $p_i = p \in (0, 1)$ for $i \in [n]$. Let $P(n, k, p)$ represent the tightest upper bound for the sum of $n$ pairwise independent identical Bernoulli random variables exceeding an integer $k \in [n]$. Then,

$$
P(n, k, p) = \begin{cases} 
1, & k < (n-1)p \quad \text{case (a)} \\
\frac{[(n-1)(1-p) + k]p}{k}, & (n-1)p \leq k < 1 + (n-1)p \quad \text{case (b)} \\
\frac{n(n-1)p^2 + (i-1)(i-2np)}{(k-i)^2 + (k-i)}, & k \geq 1 + (n-1)p, \; i = \left\lceil \frac{np(k-1) - (n-1)p}{k-np} \right\rceil \quad \text{case (c)}.
\end{cases}
$$

(3.2)

Proof. The tightest upper bound $P(n, k, p)$ is the optimal value of the linear program:

$$
P(n, k, p) = \max \sum_{c \in \{0,1\}^n : \sum_i c_i \geq k} \theta(c) \\
\text{s.t.} \sum_{c \in \{0,1\}^n} \theta(c) = 1 \\
\sum_{c \in \{0,1\}^n : c_i = 1} \theta(c) = p, \quad \forall i \in [n], \quad (3.3) \\
\sum_{c \in \{0,1\}^n : c_i = 1, c_j = 1} \theta(c) = p^2, \quad \forall (i,j) \in K_n, \\
\theta(c) \geq 0, \quad \forall c \in \{0,1\}^n,
$$

where the decision variables are the joint probabilities $\theta(c) = P(\tilde{c} = c)$ for $c \in \{0,1\}^n$. Consider the following linear program in $n+1$ variables which provides an upper bound on $P(n, k, p)$:

$$
BP(n, k, p) = \max \sum_{\ell=0}^{n} v_\ell \\
\text{s.t.} \sum_{\ell=0}^{n} v_\ell = 1 \\
\sum_{\ell=1}^{n} \ell v_\ell = np \\
\sum_{\ell=2}^{n} \binom{\ell}{2} v_\ell = \binom{n}{2} p^2 \\
v_\ell \geq 0, \quad \forall \ell \in [0,n],
$$

(3.4)

where the decision variables are the probabilities $v_\ell = P(\sum_{i=1}^{n} \tilde{c}_i = \ell)$ for $\ell \in [0,n]$. Linear programs of the form (3.4) have been studied in Boros and Prékopa [1989] in the context of aggregated binomial moment problems. As we shall see, these two formulations are equivalent with identical pairwise independent random variables.
**Step (1):** \( \mathcal{P}(n,k,p) \leq B\mathcal{P}(n,k,p) \)

Given a feasible solution to (3.3) denoted by \( \theta \), construct a feasible solution to the linear program (3.4) as:

\[
v_\ell = \sum_{c \in \{0,1\}^n : \sum_i c_i = l} \theta(c), \quad \forall \ell \in [0,n].
\]

By taking expectations on both sides of the equality (1.2), we get:

\[
\sum_{l=0}^{n} \binom{l}{j} p^n (\sum_{i=1}^{n} \hat{c}_i = l) = E[S_j(\hat{c})], \quad \forall j \in [0,n].
\]

Applying it for \( j = 0,1,2 \), we get the three equality constraints in (3.4):

\[
\sum_{\ell=0}^{n} v_\ell = 1,
\]

\[
\sum_{\ell=1}^{n} \ell v_\ell = E\left[ \sum_{i=1}^{n} \hat{c}_i \right] = np,
\]

\[
\sum_{\ell=2}^{n} \binom{\ell}{2} v_\ell = E\left[ \sum_{(i,j) \in K_n} \hat{c}_i \hat{c}_j \right] = n(n-1)p^2/2.
\]

Lastly, the objective function value of this feasible solution satisfies:

\[
\sum_{\ell=k}^{n} v_\ell = \sum_{\ell=k}^{n} \sum_{c \in \{0,1\}^n : \sum_i c_i = l} \theta(c)
\]

\[
= \sum_{c \in \{0,1\}^n : \sum_i c_i \geq k} \theta(c).
\]

Hence, \( \mathcal{P}(n,k,p) \leq B\mathcal{P}(n,k,p) \).

**Step (2):** \( \overline{\mathcal{P}}(n,k,p) \geq B\mathcal{P}(n,k,p) \)

Given an optimal solution to (3.4) denoted by \( \nu \), construct a feasible solution to the linear program (3.3) by distributing \( v_\ell \) equally among all the realizations in \( \{0,1\}^n \) with exactly \( \ell \) ones:

\[
\theta(c) = \frac{v_\ell}{\binom{n}{\ell}}, \quad \forall c \in \{0,1\}^n : \sum_{i=1}^{n} c_i = \ell, \forall \ell \in [0,n].
\]
The first constraint in (3.3) is satisfied since:

\[ \sum_{c \in \{0, 1\}^n} \theta(c) = \sum_{\ell=0}^{n} v_{\ell} \binom{n}{\ell} \]

\[ \text{[since } |\{0, 1\}^n : \sum_{i=1}^{n} c_i = \ell| = \binom{n}{\ell} \]

\[ = \sum_{\ell=0}^{n} v_{\ell} \]

\[ = 1. \]

The second constraint in (3.3) is satisfied since:

\[ \sum_{c \in \{0, 1\}^n : c_j = 1} \theta(c) = \sum_{\ell=1}^{n} \ell v_{\ell} \binom{n}{\ell} (\ell - 1) \]

\[ \text{[since } |\{0, 1\}^n : \sum_{i=1}^{n} c_i = \ell, c_j = 1| = \binom{n-1}{\ell-1} \]

\[ = \sum_{\ell=1}^{n} \frac{\ell v_{\ell}}{n} \]

\[ = p. \]

The third constraint in (3.3) satisfied since:

\[ \sum_{c \in \{0, 1\}^n : c_i = 1, c_j = 1} \theta(c) = \sum_{\ell=2}^{n} v_{\ell} \binom{n-2}{\ell-2} \]

\[ \text{[since } |\{0, 1\}^n : \sum_{t=1}^{n} c_t = \ell, c_i = 1, c_j = 1| = \binom{n-2}{\ell-2} \]

\[ = \frac{2}{n(n-1)} \sum_{\ell=2}^{n} \frac{\ell}{2} v_{\ell} \]

\[ = p^2. \]

The objective function value of the feasible solution is given by:

\[ \sum_{c \in \{0, 1\}^n : \sum_i c_i \geq k} \theta(c) = \sum_{\ell=k}^{n} \sum_{c \in \{0, 1\}^n : \sum_i c_i = \ell} \theta(c) \]

\[ = \sum_{\ell=k}^{n} v_{\ell} \]

\[ = BP(n, k, p). \]

Hence, the optimal objective value of the two linear programs are equivalent. The formula for the tight bound in the theorem is then exactly the Boros and Prekopa bound in (3.1) (the bound \(BP(n, k, p)\) is also derived in the work of Sathe et al. [1980], although tightness of the bound is not shown there). It
is also straightforward to verify that the following distributions attain the bounds for each of the cases (a)-(c) in the statement of the theorem:

(a) The probabilities are given as:

\[
\theta(c) = \begin{cases} 
\frac{(1-p)(j-(n-1)p)}{\binom{n-1}{j-1}}, & \text{if } \sum_{t=1}^{n} c_t = j-1, \\
\frac{(1-p)(1+(n-1)p-j)}{\binom{n-1}{j}}, & \text{if } \sum_{t=1}^{n} c_t = j, \\
n(n-1)p^2 + (j-1)(j-2np) \bigg/ (n-j)^2 + (n-j), & \text{if } \sum_{t=1}^{n} c_t = n,
\end{cases}
\]

where \( j = \lceil (n-1)p \rceil \) and all other support points have zero probability.

(b) The probabilities are given as:

\[
\theta(c) = \begin{cases} 
\frac{1-p}{k}(k-(n-1)p), & \text{if } \sum_{t=1}^{n} c_t = 0, \\
p(1-p) \bigg/ \binom{n-2}{k-1}, & \text{if } \sum_{t=1}^{n} c_t = k, \\
p((n-1)p-k-1) \bigg/ n-k, & \text{if } \sum_{t=1}^{n} c_t = n,
\end{cases}
\]

where all other support points have zero probability.

(c) The probabilities are given as:

\[
\theta(c) = \begin{cases} 
np[(n-1)p-(k+i-1)] + ik \bigg/ \binom{n}{i-1}(k-i+1), & \text{if } \sum_{t=1}^{n} c_t = i-1, \\
np[(k+i-2)-(n-1)p] - k(i-1) \bigg/ \binom{n}{i}(k-i), & \text{if } \sum_{t=1}^{n} c_t = i, \\
n(n-1)p^2 + (i-1)(i-2np) \bigg/ \binom{n}{i}[(k-i)^2 + (k-i)], & \text{if } \sum_{t=1}^{n} c_t = k,
\end{cases}
\]

where all other support points have zero probability and the index \( i \) is evaluated as stated in equation (3.2)(c). It is straightforward to see that with identical marginals, the tight union bound in Theorem 2.1 reduces to the bound in case (b) of Theorem 3.1.

\[ \square \]

### 3.1 Connection of Theorem 3.1 to existing results

In related work, Benjamini et al. [2012] and Peled et al. [2011] derived probability bounds (not necessarily tight) for the sum of \( t \)-wise independent Bernoulli random variables with identical probabilities (as a special case, pairwise independent random variables are studied in these papers). For the specific case, where all the random variables take a value of one (this corresponds to \( k = n \) in case (c)), the tight bound is provided in these works by making a connection with the Boros and Prekopa bound in (3.1).
The analysis in Theorem 3.1 can be easily extended to more general $t$-wise independent variables ($t \geq 3$) from the symmetry assumptions. Recent work by Garnett [2020] provides the tight upper bound on the probability that the sum of pairwise independent Bernoulli random variables with identical marginals exceeds the mean by a small amount. This corresponds to case (b). Theorem 3.1 provides the equivalence for all values of $(n,k,p)$.

3.2 Tightness of alternative bounds

We next discuss an application of Theorem 3.1. While the Boros and Prekopa bound provides the tightest upper bound with identical marginals, the formula is more involved than the Chebyshev bound which reduces to:

$$P(n,k,p) \leq \begin{cases} 1, & k < np, \\ np(1-p)/(np(1-p)+(k-np)^2), & np \leq k \leq n. \end{cases}$$

and the Schmidt, Siegel and Srinivasan bound which reduces to:

$$\bar{P}(n,k,p) \leq \min \left(1, \frac{np}{k}, \frac{n(n-1)p^2}{k(k-1)} \right).$$

It is possible to then use Theorem 3.1 to identify conditions on the parameters $(n,k,p)$ for which the bounds in (3.5) and (3.6) are tight. We only focus on the non-trivial cases where the tight bound is strictly less than one and $n \geq 3$.

**Proposition 3.1.**

(a) For $p = \alpha/(n-1)$ and any integer $\alpha \in [n-2]$, the Chebyshev bound in (3.5) is tight for the values of $k = \alpha + 1$ and $k = n$.

(b) For $p \leq 1/(n-1)$, the Schmidt, Siegel and Srinivasan bound in (3.6) is tight for all $k \in [2,n]$ while for $p > 1/(n-1)$, the bound is tight for all $k \in \left[\lceil 1 + (n-1)p \rceil, \lfloor n(n-1)p^2/(np-1) \rfloor \right]$.

**Proof.** Since Theorem 3.1 provides the tight bound, we simply need to show the equivalence with the bounds in (3.5) and (3.6) for the instances in the proposition.

(a) Consider $p = \alpha/(n-1)$ for any integer $\alpha \in [n-2]$.

1. Set $k = \alpha + 1$. This corresponds to case (c) in Theorem 3.1. Plugging in the values, the index $i$ which is required for finding the tight bound is given by:

$$i = \left\lceil \frac{na(\alpha + 1 - \alpha)/(n-1)}{\alpha + 1 - na/(n-1)} \right\rceil = 0.$$

The corresponding tight bound in (3.2) gives:

$$P(n,k,p) = \frac{na}{(n-1)(\alpha + 1)} = \frac{np}{np + 1 - p}.$$
It is straightforward to verify by plugging in the values that the Chebyshev bound is exactly the same.

2. Set $k = n$. This corresponds to case (c) in Theorem 3.1. Plugging in the values, the index $i$ in the tight bound is given by:

$$i = \left\lceil \frac{n\alpha(n - 1 - \alpha)/(n - 1)}{n - n\alpha/(n - 1)} \right\rceil = \alpha.$$ 

The tight bound in (3.2) gives:

$$\mathcal{P}(n, k, p) = \frac{\alpha}{(n - 1)(n - \alpha)} \frac{p}{p + n(1 - p)}.$$ 

It is straightforward to verify by plugging in the values that the Chebyshev bound is exactly the same in this case.

(b) Observe that the last two terms in the Schmidt, Siegel and Srinivasan bound in (3.6) satisfy:

$$\frac{n(n - 1)p^2}{k(k - 1)} \leq \frac{np}{k} \text{ when } k \geq 1 + (n - 1)p.$$ 

Since this implies $1 \geq np/k$, the bound in (3.6) reduces to $n(n-1)p^2/k(k-1)$. The range of $k \geq 1 + (n - 1)p$ corresponds to case (c) in Theorem 3.1. If $k = 1 + (n - 1)p$, the index $i = \left\lfloor \frac{np(k - (1 + (n - 1)p))}{k - np} \right\rfloor = 0$ and the tight bound is:

$$\mathcal{P}(n, k, p) = \frac{np}{1 + (n - 1)p},$$

which is exactly the Schmidt, Siegel and Srinivasan bound. We can also verify that when the index $i = 1$ in case (c), then the tight bound in Theorem 3.1 reduces to:

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

We now identify conditions when $k > 1 + (n - 1)p$ and the index $i$ is equal to one.

1. Set $0 < p < 1/(n - 1)$. For the values of the $p$ in this interval, the valid range of $k$ in case (c) corresponds to all integer values of $k > 1 + (n - 1)p$ which means $k \geq 2$. For the probability
\[ 0 < p \leq 1/n, \text{ the index } i \text{ satisfies:} \]
\[
i = \left\lceil \frac{np(k - np - (1 - p))}{k - np} \right\rceil
\]
\[
= \left\lceil np\left(1 - \frac{1 - p}{k - np}\right)\right\rceil
\]
\[
= 1 \quad \text{[since } 0 < np \leq 1 \text{ and } (1 - p) \in (0, 1) \text{ and } k - np \geq 1].
\]

For the probability \(1/n < p < 1/(n - 1)\), let \((n - 1)p = 1 - \delta\) where \(\delta < 1\). Then, since \(np > 1\), we have \(\frac{(1 - \delta)}{n - 1} > 1\) or equivalently \(n\delta < 1\). The index \(i\) satisfies:
\[
i = \left\lceil \frac{np((n - 1)p - (k - 1))}{np - k} \right\rceil
\]
\[
< \left\lceil \frac{np(1 - \delta - (k - 1))}{1 - k} \right\rceil
\]
\[
[\text{since } np > 1 \text{ and } (n - 1)p = 1 - \delta]
\]
\[
= \left\lceil \frac{np(k - 2 + \delta)}{k - 1} \right\rceil
\]
\[
< \left\lceil \frac{n(k - 2 + \delta)}{(n - 1)(k - 1)} \right\rceil
\]
\[
[\text{since } p < 1/(n - 1)]
\]
\[
= \left\lceil \frac{n(k - 2 + \delta)}{nk - n - k + 1} \right\rceil
\]
\[
\leq \left\lceil \frac{n(k - 2 + \delta)}{nk - 2n + 1} \right\rceil
\]
\[
[\text{since } k \leq n]
\]
\[
= \left\lceil \frac{n(k - 2) + n\delta}{n(k - 2) + 1} \right\rceil
\]
\[
= 1 \quad \text{[since } k \geq 2 \text{ and } 0 < n\delta < 1]\]

Hence, the bound in (3.6) is tight in this case for all integer values of \(k \geq 2\).

2. For \(p > 1/(n - 1)\), the index \(i = 1\) when \(k(np - 1) \leq n(n - 1)p^2\). This corresponds to all integer values \(k \in \left\lceil 1 + (n - 1)p \right\rceil, \left\lfloor n(n - 1)p^2/(np - 1) \right\rfloor\).

\[\square\]

A specific instance to show the tightness of the Chebyshev bound is to set \(p = 1/2, k = n\) and \(n = 2^m - 1\) using \(m\) independent Bernoulli random variables to construct \(n\) pairwise independent Bernoulli
random variables (see Tao [2012], Goemans [2015], Pass and Spektor [2018] for this construction). Proposition 3.1(a) includes this instance (set $\alpha = (n - 1)/2$, $k = n$ and $n = 2^m - 1$). In addition, Proposition 3.1(a) identifies other values of $p$ and $k$ where the Chebyshev bound is tight. Proposition 3.1(b) also shows that the Schmidt, Siegel and Srinivasan bound is tight for identical marginals for small probability values ($p \leq 1/(n - 1)$), for all values of $k$, except $k = 1$. We now provide a numerical illustration of the results in Theorem 3.1 and Proposition 3.1.

**Example 3 (Identical marginals).** In Table 3, we provide a numerical comparison of the bounds for $n = 11$ for a set of values of $p$ and $k$. The conditions identified in Proposition 3.1 covers all the instances in Table 3 where the Chebyshev bound and the Schmidt, Siegel and Srinivasan bound are tight. The instances when the Chebyshev bound is tight correspond to (i) $p = 0.1$ (here $\alpha = 1$ and the Chebyshev bound is tight for $k = 2$ and $k = 10$), (ii) $p = 0.2$ (here $\alpha = 2$ and the Chebyshev bound is tight for $k = 3$ and $k = 10$) and (iii) $p = 0.5$ (here $\alpha = 5$ and the Chebyshev bound is tight for $k = 6$ and $k = 10$). The Schmidt, Siegel and Srinivasan bound is tight for the small values of $p = 0.01, 0.05, 0.10$ (which are less than or equal to $1/(n - 1) = 0.1$) and for all values of $k$, except $k = 1$.

Table 3: Upper bound on probability of sum of random variables for $n = 11$. For each value of $p$ and $k$, the table provides the tight bound in (3.2) followed by the Chebyshev bound (3.5) and the Schmidt, Siegel and Srinivasan bound (3.6). The underlined instances illustrate cases when the upper bounds in either (3.5) or (3.6) are tight.

<table>
<thead>
<tr>
<th>$p/k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.10900</td>
<td>0.05500</td>
<td>0.0184</td>
<td>0.0092</td>
<td>0.0065</td>
<td>0.0037</td>
<td>0.0027</td>
<td>0.0020</td>
<td>0.0016</td>
<td>0.0013</td>
<td>0.00010</td>
</tr>
<tr>
<td>0.12087</td>
<td>0.02959</td>
<td>0.01288</td>
<td>0.00715</td>
<td>0.00454</td>
<td>0.00313</td>
<td>0.00229</td>
<td>0.00175</td>
<td>0.00138</td>
<td>0.00112</td>
<td>0.00092</td>
<td></td>
</tr>
<tr>
<td>0.11000</td>
<td>0.00550</td>
<td>0.00184</td>
<td>0.00092</td>
<td>0.00055</td>
<td>0.00037</td>
<td>0.00027</td>
<td>0.00020</td>
<td>0.00016</td>
<td>0.00013</td>
<td>0.00010</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>0.52500</td>
<td>0.13750</td>
<td>0.04583</td>
<td>0.02292</td>
<td>0.01375</td>
<td>0.00917</td>
<td>0.00655</td>
<td>0.00491</td>
<td>0.00382</td>
<td>0.00306</td>
<td>0.00250</td>
</tr>
<tr>
<td>0.72069</td>
<td>0.19905</td>
<td>0.08008</td>
<td>0.04205</td>
<td>0.02571</td>
<td>0.01729</td>
<td>0.01240</td>
<td>0.00933</td>
<td>0.00726</td>
<td>0.00582</td>
<td>0.00477</td>
<td></td>
</tr>
<tr>
<td>0.55000</td>
<td>0.13750</td>
<td>0.04583</td>
<td>0.02292</td>
<td>0.01375</td>
<td>0.00917</td>
<td>0.00655</td>
<td>0.00491</td>
<td>0.00382</td>
<td>0.00306</td>
<td>0.00250</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>1</td>
<td>0.55000</td>
<td>0.18333</td>
<td>0.09167</td>
<td>0.05500</td>
<td>0.03667</td>
<td>0.02620</td>
<td>0.01965</td>
<td>0.01528</td>
<td>0.01223</td>
<td>0.01000</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.55000</td>
<td>0.21522</td>
<td>0.10532</td>
<td>0.06112</td>
<td>0.03960</td>
<td>0.02766</td>
<td>0.02038</td>
<td>0.01562</td>
<td>0.01235</td>
<td>0.01000</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.55000</td>
<td>0.18333</td>
<td>0.09167</td>
<td>0.05500</td>
<td>0.03667</td>
<td>0.02620</td>
<td>0.01965</td>
<td>0.01528</td>
<td>0.01233</td>
<td>0.01000</td>
</tr>
<tr>
<td>0.11</td>
<td>1</td>
<td>0.59950</td>
<td>0.22184</td>
<td>0.11092</td>
<td>0.06665</td>
<td>0.04437</td>
<td>0.03037</td>
<td>0.02170</td>
<td>0.01627</td>
<td>0.01266</td>
<td>0.01013</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.63310</td>
<td>0.25156</td>
<td>0.12154</td>
<td>0.06975</td>
<td>0.04484</td>
<td>0.03133</td>
<td>0.02283</td>
<td>0.01744</td>
<td>0.01375</td>
<td>0.01112</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.65000</td>
<td>0.22184</td>
<td>0.11092</td>
<td>0.06665</td>
<td>0.04437</td>
<td>0.03170</td>
<td>0.02377</td>
<td>0.01849</td>
<td>0.01479</td>
<td>0.01210</td>
</tr>
<tr>
<td>0.15</td>
<td>1</td>
<td>0.78750</td>
<td>0.41250</td>
<td>0.19584</td>
<td>0.09792</td>
<td>0.05875</td>
<td>0.039167</td>
<td>0.02798</td>
<td>0.020983</td>
<td>0.01632</td>
<td>0.01306</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.91968</td>
<td>0.43489</td>
<td>0.20253</td>
<td>0.11109</td>
<td>0.06901</td>
<td>0.04672</td>
<td>0.03362</td>
<td>0.02531</td>
<td>0.01972</td>
<td>0.01579</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.82500</td>
<td>0.41250</td>
<td>0.20625</td>
<td>0.12375</td>
<td>0.08250</td>
<td>0.05893</td>
<td>0.044197</td>
<td>0.034375</td>
<td>0.02750</td>
<td>0.02250</td>
</tr>
<tr>
<td>0.20</td>
<td>1</td>
<td>1</td>
<td>0.73334</td>
<td>0.33334</td>
<td>0.16667</td>
<td>0.10000</td>
<td>0.06667</td>
<td>0.04762</td>
<td>0.03572</td>
<td>0.02778</td>
<td>0.02223</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.73334</td>
<td>0.35200</td>
<td>0.18324</td>
<td>0.10865</td>
<td>0.07097</td>
<td>0.04972</td>
<td>0.03667</td>
<td>0.02812</td>
<td>0.02223</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.73334</td>
<td>0.36667</td>
<td>0.22000</td>
<td>0.14667</td>
<td>0.10477</td>
<td>0.07858</td>
<td>0.06112</td>
<td>0.04889</td>
<td>0.04000</td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.91667</td>
<td>0.54167</td>
<td>0.29167</td>
<td>0.17500</td>
<td>0.11667</td>
<td>0.08334</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.91667</td>
<td>0.55000</td>
<td>0.30556</td>
<td>0.18334</td>
<td>0.11957</td>
<td>0.08334</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.91667</td>
<td>0.65477</td>
<td>0.49108</td>
<td>0.38195</td>
<td>0.30556</td>
<td>0.25000</td>
<td></td>
</tr>
</tbody>
</table>

It is also clear why the Schmidt, Siegel and Srinivasan bound is not tight for $k = 1$, since it just reduces to the Markov bound $np$ and does not exploit the pairwise independence information. For $k = 1$, the tight bound from Theorem 3.1 is given by $np - (n - 1)p^2$ (see Theorem 2.1 which reduces to the same bound for $k = 1$). For larger values of $p$ above 0.1, such as $p = 0.11$ in the table, from Proposition 3.1(b), the Schmidt, Siegel and Srinivasan bound is tight for $k \in [[2.1, [6.33]]$ which corresponds to
4 Improved bounds with non-identical marginals

The next theorem provides new probability bounds for the sum of pairwise independent random variables with possibly non-identical marginals when \( k \geq 2 \). These bounds build on the previous results while exploiting the ordering of probabilities for the pairwise independent random variables.

**Theorem 4.1.** Sort the input probabilities in increasing order as \( p_1 \leq p_2 \leq \ldots \leq p_n \). Define the partial binomial moment \( S_{1r} = \sum_{i=1}^{n-r} p_i \) for \( r \in [0, n-1] \) and \( S_{2r} = \sum_{(i,j) \in K_{n-r}} p_i p_j \) for \( r \in [0, n-2] \).

(a) The ordered Schmidt, Siegel and Srinivasan bound is a valid upper bound on \( \overline{P}(n, k, p) \):

\[
\overline{P}(n, k, p) \leq \min \left( 1, \min_{0 \leq r_1 \leq k-1} \frac{S_{1r_1}}{k - r_1}, \min_{0 \leq r_2 \leq k-2} \frac{S_{2r_2}}{(k-r_2)^2} \right), \quad \forall k \in [2, n],
\]

(b) The ordered Boros and Prekopa bound is a valid upper bound on \( \overline{P}(n, k, p) \):

\[
\overline{P}(n, k, p) \leq \min_{0 \leq r \leq k-1} BP(n-r, k-r, p), \quad \forall k \in [2, n],
\]

where:

\[
BP(n-r, k-r, p) = \begin{cases} 
1, & k < \frac{(n-r-1)S_{1r} - 2S_{2r}}{n-r - S_{1r}} + r, \\
\frac{(k-r+n-r-1)S_{1r} - 2S_{2r}}{(k-r)(n-r)}, & \frac{(n-r-1)S_{1r} - 2S_{2r}}{n-r - S_{1r}} + r \leq k < 1 + \frac{2S_{2r}}{S_{1r}} + r, \\
\frac{(i-1)(i-2S_{1r}) + 2S_{2r}}{(k-r-i)^2 + (k-r-i)}, & k \geq 1 + \frac{2S_{2r}}{S_{1r}} + r, \quad i = \left\lceil \frac{(k-r-1)S_{1r} - 2S_{2r}}{k-r - S_{1r}} \right\rceil.
\end{cases}
\]

(c) The ordered Chebyshev bound is a valid upper bound on \( \overline{P}(n, k, p) \):

\[
\overline{P}(n, k, p) \leq \min_{0 \leq r \leq k-1} CH(n-r, k-r, p), \quad \forall k \in [2, n],
\]

where:

\[
CH(n-r, k-r, p) = \begin{cases} 
1, & k < S_{1r} + r, \\
\frac{S_{1r} - (S_{1r}^2 - 2S_{2r})}{S_{1r} - (S_{1r}^2 - 2S_{2r}) + (k-r-S_{1r})^2}, & S_{1r} + r \leq k \leq n.
\end{cases}
\]

**Proof.**

\( k \in [3, 6] \). This can be similarly verified for the other probabilities \( p = 0.15, 0.2, 0.5 \) in the table.
(a) We observe that for any $0 \leq r_1 \leq k - 1$ and any subset $S \subseteq [n]$ of the random variables of cardinality $n - r_1$, an upper bound is given as:

$$
\mathbb{P}
\left(
\sum_{i=1}^{n} \tilde{c}_i \geq k
\right) \leq \mathbb{P}
\left(
\sum_{i \in S} \tilde{c}_i \geq k - r_1
\right)
$$

[since $\sum_{i=1}^{n} c_i \geq k$ for $c \in \{0,1\}^n$ implies $\sum_{i \in S} c_i \geq k - r_1$ for $c \in \{0,1\}^n$]

$$
\leq \frac{\mathbb{E} \left[ \sum_{i \in S} \tilde{c}_i \right]}{k - r_1}
$$

[using Markov’s inequality]

$$
= \sum_{i \in S} p_i
$$

The tightest upper bound of this form is obtained by minimizing over all $0 \leq r_1 \leq k - 1$ and subsets $S \subseteq [n]$ with $|S| = n - r_1$, which gives:

$$
\mathbb{P}
\left(
\sum_{i=1}^{n} \tilde{c}_i \geq k
\right) \leq \min_{0 \leq r_1 \leq k - 1} \min_{S: |S| = n - r_1} \sum_{i \in S} p_i
$$

$$
= \min_{0 \leq r_1 \leq k - 1} \sum_{i=1}^{n-r_1} p_i
$$

[using the $n-r_1$ smallest probabilities].

We derive the next term in (4.1) using a similar approach while accounting for pairwise independence. For any $0 \leq r_2 \leq k - 2$ and any subset $S \subseteq [n]$ of the random variables of cardinality $n - r_2$, an upper bound is given by:

$$
\mathbb{P}
\left(
\sum_{i=1}^{n} \tilde{c}_i \geq k
\right) \leq \mathbb{P}
\left(
\sum_{i \in S} \tilde{c}_i \geq k - r_2
\right)
$$

$$
= \mathbb{P}
\left(
\begin{pmatrix}
\sum_{i \in S} \tilde{c}_i \\
2
\end{pmatrix}
\geq
\begin{pmatrix}
(k - r_2) \\
2
\end{pmatrix}
\right)
$$

$$
\leq \mathbb{E}
\left[
\begin{pmatrix}
\sum_{i \in S} \sum_{j \in S: j > i} \tilde{c}_i \tilde{c}_j \\
2
\end{pmatrix}
\right]
$$

[using equation (1.2) and Markov’s inequality]

$$
= \frac{\sum_{i \in S} \sum_{j \in S: j > i} \mathbb{E}[\tilde{c}_i] \mathbb{E}[\tilde{c}_j]}{(k - r_2)_2}
$$

[using pairwise independence]

$$
= \frac{\sum_{i \in S} \sum_{j \in S: j > i} p_i p_j}{(k - r_2)_2}
$$

The tightest upper bound of this form is obtained by minimizing over $0 \leq r_2 \leq k - 2$ and all sets.
\( S \) of size \( n - r_2 \). This gives:
\[
\mathbb{P}\left( \sum_{i=1}^{n} \tilde{c}_i \geq k \right) \leq \min_{0 \leq r_2 \leq k-2} \min_{S:|S|=n-r_2} \frac{\sum_{i \in S} \sum_{j \in S: j > i} p_i p_j}{(k-r_2)^2} \\
= \min_{0 \leq r_2 \leq k-2} \left( \frac{\sum_{(i,j) \in K_{n-r_2}} p_i p_j}{(k-r_2)^2} \right) \tag{4.5}
\]
[using the \( n - r_2 \) smallest probabilities].

From the bounds (4.4) and (4.5), we get:
\[
\overline{P}(n, k, p) \leq \min \left( 1, \min_{0 \leq r_1 \leq k-1} \left( \frac{S_{1r_1}}{k-r_1} \right), \min_{0 \leq r_2 \leq k-2} \left( \frac{S_{2r_2}}{(k-r_2)^2} \right) \right), \quad \forall k \in [2, n]
\]
where \( S_{1r_1} = \sum_{i=1}^{n-r_1} p_i \) for \( r_1 \in [0, n-1] \) and \( S_{2r_2} = \sum_{(i,j) \in K_{n-r_2}} p_i p_j \) for \( r_2 \in [0, n-2] \). It is straightforward to see that this approach is essentially creating a set of dual feasible solutions and picking the best among it. The dual formulation is:
\[
\overline{P}(n, k, p) = \min \sum_{(i,j) \in K_n} \lambda_{ij} p_i p_j + \sum_{i=1}^{n} \lambda_i p_i + \lambda_0 \\
\text{s.t} \quad \sum_{(i,j) \in K_n} \lambda_{ij} c_i c_j + \sum_{i=1}^{n} \lambda_i c_i + \lambda_0 \geq 0, \quad \forall c \in \{0, 1\}^n \\
\sum_{(i,j) \in K_n} \lambda_{ij} c_i c_j + \sum_{i=1}^{n} \lambda_i c_i + \lambda_0 \geq 1, \quad \forall c \in \{0, 1\}^n : \sum_t c_t \geq k.
\]
Each component of the second term is obtained by choosing dual feasible solutions with \( \lambda_i = 1/(k-r_1) \) for \( i \in [n-r_1] \) and setting all other dual variables to 0. Similarly, each component of the third term is obtained by choosing dual feasible solutions with \( \lambda_{ij} = 1/(k-r_2^2) \) for \( (i,j) \in K_{n-r_2} \) and setting all other dual variables to 0.

(b) The bound in (4.2) is obtained by using the inequality:
\[
\mathbb{P}\left( \sum_{i=1}^{n} \tilde{c}_i \geq k \right) \leq \mathbb{P}\left( \sum_{i=1}^{n-r} \tilde{c}_i \geq k-r \right), \quad \forall r \in [0, k-1].
\]
Then, we compute an upper bound on \( \mathbb{P}\left( \sum_{i=1}^{n-r} \tilde{c}_i \geq k-r \right) \) by using the aggregated moments \( S_{1r} \).
and $S_{2r}$ with the Boros and Prekopa bound from (3.1) as follows:

$$BP(n-r,k-r,p) = \begin{cases} 1, & k < \frac{(n-r-1)S_{1r} - 2S_{2r}}{n-r-S_{1r}} + r \\ \frac{(k-r + n-r-1)S_{1r} - 2S_{2r}}{(k-r)(n-r)}, & \frac{(n-r-1)S_{1r} - 2S_{2r}}{n-r-S_{1r}} + r \leq k \leq 1 + \frac{2S_{2r}}{S_{1r}} + r \\ \frac{(i-1)(i-2S_{1r}) + 2S_{2r}}{(k-r-i)^2 + (k-r-i)}, & k \geq 1 + \frac{2S_{2r}}{S_{1r}} + r, \ i = \left\lceil \frac{(k-r-1)S_{1r} - 2S_{2r}}{k-r-S_{1r}} \right\rceil \end{cases}$$

Since the relation $P(n,k,p) \leq BP(n-r,k-r,p)$ is satisfied for every $0 \leq r \leq k-1$, the upper bound on $P(n,k,p)$ is obtained by taking the minimum over all possible values of $r$:

$$P(n,k,p) \leq \min_{0 \leq r \leq k-1} BP(n-r,k-r,p).$$

(c) Proceeding in a similar manner as in (b), by using the aggregated moments $S_{1r}$ and $S_{2r}$ with Chebyshev bound, the upper bound for a given $r$ ($0 \leq r \leq k-1$) can be written as follows:

$$CH(n-r, k-r, p) = \begin{cases} 1, & k < S_{1r} + r \\ \frac{S_{1r} - (S_{2r}^2 - 2S_{2r})}{S_{1r} - (S_{2r}^2 - 2S_{2r}) + (k-r-S_{1r})^2}, & S_{1r} + r \leq k \leq n. \end{cases}$$

The upper bound on $\bar{P}(n,k,p)$ is obtained by taking the minimum over all possible values of $r$:

$$\bar{P}(n,k,p) \leq \min_{0 \leq r \leq k-1} CH(n-r,k-r,p), \ \forall k \in [2,n]$$

4.1 Connection to existing results

Prior work in Rüger [1978] shows that ordering of probabilities provides the tightest upper bound on the probability of the sum of Bernoulli random variables exceeding $k$ while allowing for arbitrary dependence. Specifically, the bound derived there is:

$$\min\left(1, \min_{0 \leq r \leq k-1} \left(\frac{S_{1r}}{k-r}\right)\right).$$

However, this bound does not use pairwise independence information. Part (a) of Theorem 4.1 tightens the analysis in Rüger [1978] for pairwise independent random variables. It is also straightforward to see that the ordered Schmidt, Siegel and Srinivasan bound in (4.1) is at least as good as the bound in (1.3) (simply plug in $r = 0$). Building on the ordering of probabilities, the bound in (4.2) uses aggregated binomial moments for $k$ ordered sets of random variables of size $n-r$ where $0 \leq r \leq k-1$. When $r = 0$, the bound in (4.2) reduces to the original aggregated moment bound of Boros and Prekopa in (3.1) and hence this bound is at least as tight. Further, the bounds in Theorem 4.1 are clearly efficiently
computable. We next provide two numerical examples to illustrate the impact of ordering on the quality of the three bounds.

**Example 4 (Non-identical marginals).** Consider an example with $n = 12$ random variables with the probabilities given by

$$p = (0.0651, 0.0977, 0.1220, 0.1705, 0.3046, 0.4402, 0.4952, 0.6075, 0.6842, 0.8084, 0.9489, 0.9656).$$

Table 4 compares the three ordered bounds with the three unordered bounds and the corresponding tight bound. Numerically, the ordered Boros and Prekopa bound is found to be tight in this example for $k = 7, 8, 9, 12$ while the ordered Schmidt, Siegel and Srinivasan bound is tight for $k = 12$. The Boros and Prekopa bound is uniformly the best performing of the three bounds, while among the other two bounds, none uniformly dominates the other. For example, comparing the ordered bounds when $7 \leq k \leq 9$, the Chebyshev bound outperforms the Schmidt, Siegel and Srinivasan bound, but when $k = 6$ or $10 \leq k \leq 12$, the Schmidt, Siegel and Srinivasan bound does better. Comparing the unordered bounds when $7 \leq k \leq 9$, the Schmidt, Siegel and Srinivasan bound outperforms the Chebyshev bound when $k = 6$ but for all $k \geq 7$, the Chebyshev bound does better. In terms of absolute difference between ordered and unordered bounds, ordering appears to provide the maximum improvement to the Schmidt, Siegel and Srinivasan bound, followed by the Boros and Prekopa and the Chebyshev bound.

**Example 5 (Non-identical marginals).** In this example, we numerically compute the improvement of the new ordered bounds over the unordered bounds for $n = 100$ variables by creating 500 instances by randomly generating the probabilities $p = (p_1, p_2, ..., p_{100})$. First, we consider small marginal probabilities by uniformly and independently generating the entries of $p$ between 0.01 and 0.05. When $k = n$, Figure 4a plots the three ordered bounds while Figure 4b shows the percentage improvement of the three bounds over their unordered counterparts. The percentage improvement is computed as ([unordered-ordered]/unordered) × 100%. In this example with small marginals, the ordered Schmidt, Siegel and Srinivasan bound is equal to the ordered Boros and Prekopa bound as seen in Figure 4a. Ordering tends to improve the Schmidt, Siegel and Srinivasan bound significantly for smaller probabilities, since both
the partial binomial moment terms $S_{1r}$ and $S_{2r}$ are smaller with smaller marginal probabilities for all $r \in [0, k - 1]$.

Figure 4: Smaller marginal probabilities $p_i$ with $n = 100, k = 100$ and 500 instances.

The percentage improvement due to ordering in figure 4b is consistently above 80% for the Schmidt, Siegel and Srinivasan bound, being while that of the Boros and Prekopa bound is around 60%. The ordered Chebyshev bound shows an almost negligible improvement by ordering in this example.

Next, we consider similar plots when $k = n - 1$ with larger marginal probabilities. The entries of $p$ are generated uniformly and independently between 0.05 and 0.99.

Figure 5: Larger marginal probabilities $p_i$ with $n = 100, k = 99$ and 500 instances.

In Figure 5a, the ordered Chebyshev bound from (4.3) performs better than the ordered Schmidt, Siegel and Srinivasan bound from (4.1). In Figure 5b, the percentage improvement due to ordering is again most significant for the Schmidt, Siegel and Srinivasan bound, being consistently above 90% while
that of the Boros and Prekopa bound is less than 40% and that of the Chebyshev bound is less than 20%. It is also clear from Figures 4 and 5 that the ordered Boros and Prekopa bound from (4.2) is the tightest of the three bounds across the instances, while among the other two bounds, none uniformly dominates the other.

4.2 Tightness in a special case

While the ordered bounds in Theorem 4.1 are not tight in general, the next proposition identifies a special case with almost identical marginals where the bound of Schmidt, Siegel and Srinivasan in (4.1) and Boros and Prekopa in (4.2) are shown to be attained.

**Proposition 4.1.** Suppose the marginal probabilities equal \( p \in (0, 1/(n-1)) \) for \( n-1 \) random variables and \( q \in (0, 1) \) for one random variable. Then, the bounds in (4.1) and (4.2) are tight for the following three instances and given by the bound:

\[
\mathcal{P}(n, k, p, q) = \begin{cases} 
\binom{n-1}{k-1}p^2, & k \geq 3, \; q \geq (n-2)p \quad \text{case (a)} \\
\binom{n-1}{k-1}p^2, & k \in [\lceil 2 + (n-2)p/q \rceil, \; n], \; p \leq q < (n-2)p \quad \text{case (b)} \\
pq, & k = n, \; 0 < q < p \quad \text{case (c)}
\end{cases}
\]

(4.6)

**Proof.** We first prove that the ordered bounds of Schmidt, Siegel and Srinivasan and Boros and Prekopa reduce to the bound in (4.6) in each of the three cases and then show that the bound is tight.

**Step (1): Show reduction of ordered bounds to the bound in (4.6)**

Let \( \mathcal{P}(n, k, p, q) \) represent the tightest upper bound when \( n-1 \) probabilities are \( p \) and one is \( q \). It can be observed that the bound in (4.6) is non-trivial for the three instances since:

\[
\binom{n-1}{k-1}p^2 = \frac{(n-1)p(n-2)p}{(k-1)(k-2)} < 1
\]

\[
[\text{since } (n-2)p < (n-1)p \leq 1 \text{ and } k \geq 3 \text{ for cases (a) and (b)},]
\]

\[
pq < 1
\]

\[
[\text{since } q < p < 1 \text{ for case (c)}].
\]

It is easy to verify that the ordered Schmidt, Siegel and Srinivasan bound in (4.1) reduces to the bound in (4.6) for a specific parameter \( r_2 \) in each of the three cases:

\[
r_2 = 1, \quad \text{cases (a) and (b)}
\]

\[
r_2 = n-2, \quad \text{case (c)}.
\]

(4.7)

It can be similarly verified that the ordered Boros and Prekopa bound in (4.2) reduces to the bound in
Equation (4.6) with the following parameters $r$ and $i$ in each of the three cases:

\[ r = 1, \ i = 0 \quad \text{cases (a) and (b)} \]
\[ r = n - 2, \ i = 0 \quad \text{case (c)} \]  

(4.8)

The effectiveness of ordering is demonstrated by (4.7) and (4.8) in that the ordered bounds of Schmidt, Siegel and Srinivasan and Boros and Prekopa correspond to $r > 0$ while their unordered counterparts in (1.3) and (3.1) correspond to $r = 0$ (considering all $n$ variables). The unordered bounds are thus strictly weaker than the ordered bounds which in turn are tight as proved in the next step.

**Step (2): Prove tightness of the bound in (4.6) by constructing extremal distributions**

Consider the linear program to compute $P(n, k, p, q)$ which can be written as:

\[
\begin{align*}
\overline{P}(n, k, p, q) &= \max \sum_{c \in \{0,1\}^n : \sum c_t \geq k} \theta(c) \\
\text{s.t} \quad &\sum_{c \in \{0,1\}^n} \theta(c) = 1 \\
&\sum_{c \in \{0,1\}^n : c_t = 1} \theta(c) = p, \quad \forall i \in [n] \\
&\sum_{c \in \{0,1\}^n : c_n = 1} \theta(c) = q \\
&\sum_{c \in \{0,1\}^n : c_i = 1, c_j = 1} \theta(c) = p^2, \quad \forall (i, j) \in K_{n-1} \\
&\sum_{c \in \{0,1\}^n : c_i = 1, c_{n-1} = 1} \theta(c) = pq, \quad \forall i \in [n-1] \\
&\theta(c) \geq 0, \quad \forall c \in \{0,1\}^n.
\end{align*}
\]

(4.9)

We now proceed to prove tightness of the bound in (4.6) for each of the three instances of the $(n, k, p, q)$ tuple by constructing feasible distributions of (4.9) which attain the bound.

(a) $P(n, k, p, q) = \frac{(n-1)p^2}{k^2}$ (cases (a) and (b)):

The following distribution attains the tight bound:

\[
\theta(c) = \begin{cases} 
(1 - q)(1 - (n-1)p), & \text{if } \sum_{t=1}^{n-1} c_t = 0 \\
 p(1 - q), & \text{if } \sum_{t=1}^{n-1} c_t = 1, c_n = 0 \\
 q(1 - (n-1)p) + \frac{(n-1)(n-2)p^2}{k(k-1)}, & \text{if } \sum_{t=1}^{n-1} c_t = 0, c_n = 1 \\
 p(q - \frac{n-2}{k-2}p), & \text{if } \sum_{t=1}^{n-1} c_t = 1, c_n = 1 \\
\frac{p^2}{(k-3)}, & \text{if } \sum_{t=1}^{n-1} c_t = k - 1, c_n = 1 
\end{cases}
\]

(4.10)
We use symbols $x, y, z, u, v$ to denote the probability of the associated scenarios in (4.10). The constraints in (4.9) reduce to:

\[
\begin{align*}
{n-3 \choose k-3} v + p^2 &= \frac{(n-3)!}{(k-3)!} p^2 \\
{n-2 \choose k-2} v + u &= \frac{(n-2)!}{(k-2)!} u + v = 1 \quad \text{and using } x,y,z,u,v \text{ from (4.10), it can be easily verified that all of the above constraints are satisfied. The non-negativity constraints for } y,v \text{ are satisfied while } x \geq 0, z \geq 0 \text{ is satisfied since } (n-1)p \leq 1. \text{ Remaining case is } u, \text{ for which we have:}
\end{align*}
\]

\[
\begin{align*}
\text{case (a): } u &= p(q - \frac{n-2}{k-2}p) \\
&\geq y = p(q - \frac{n-2}{k-2}p) \\
&\quad \text{[since } k \geq 3] \\
&= p(q - (n-2)p) \\
&\quad \text{[since } q > (n-2)p] \\
&\geq 0 \\
\text{case (b): } u &= p(q - \frac{n-2}{k-2}p) \\
&\geq p(q - \frac{n-2}{k-2}q) \\
&\quad \text{[since } k \geq 2 + (n-2)p/q] \\
&= 0.
\end{align*}
\]

The only support points contributing to the objective function are the first set of $\binom{n-1}{k-1}$ scenarios, and so we have $P(n, k, p, q) = \binom{n-1}{k-1} p^2 = \frac{(n-1)! p^2}{(k-1)!}.$

(b) $P(n, k, p, q) = pq$ (case (c)):

The following distribution attains the tight bound $pq$:

\[
\theta(c) = \begin{cases} 
(1-p)(1-(n-2)p-q), & \text{if } \sum_{t=1}^{n} c_t = 0 \quad (x) \\
p(1-p), & \text{if } \sum_{t=1}^{n} c_t = 1, c_n = 0 \quad (y) \\
q(1-p), & \text{if } \sum_{t=1}^{n} c_t = 0, c_n = 1 \quad (z) \quad (4.11) \\
p(p-q), & \text{if } \sum_{t=1}^{n} c_t = n - 1, c_n = 0 \quad (u) \\
pq, & \text{if } \sum_{t=1}^{n} c_t = n \quad (v)
\end{cases}
\]
The constraints in (4.9) reduce to:

\begin{align*}
y + u + v &= p \\
z + v &= q \\
u + v &= p^2 \\
v &= pq \\
x + y + z + u + v &= 1
\end{align*}

and using \(x, y, z, u, v\) from (4.11), it can be easily verified that all of the above constraints are satisfied. The non-negativity constraints for \(y, z, u, v\) are satisfied by \(0 < q \leq p \leq 1\) while for \(x\), we have:

\begin{align*}
x &= (1 - p)(1 - (n - 2)p - q) \\
&\geq (1 - p)(1 - (n - 2)p - p) \\
&\quad \text{[since } q < p]\ \\
&= (1 - p)(1 - (n - 1)p) \\
&\geq 0 \\
&\quad \text{[since } (n - 1)p \leq 1]\end{align*}

The distribution in (4.11) attains the bound \(pq\). We have thus constructed two feasible probability distributions in (4.10) and (4.11) which attain the bound in (4.6) in each of the three instances defined by the \((n, k, p, q)\) tuple. Hence the parameters \(r_2\), \(r\) in (4.7) and (4.8) defined for each of the three cases must be the minimizers which exactly reduce the ordered bounds in (4.1) and (4.2) to the tight bound in (4.6).

**Example 6.** This example demonstrates the usefulness of proposition 4.1 when \(n = 100\) and \(p = 0.01\) \((n - 1)p \leq 1\). It compares the tight bounds computed from (4.6) with the unordered bounds of Schmidt, Siegel and Srinivasan from (1.3) and that of Boros and Prekopa from (3.1).

![Figure 6: Comparison of unordered bounds with tight bound when n = 100, p = 0.01](image)

Figure 6: Comparison of unordered bounds with tight bound when \(n = 100\), \(p = 0.01\)
Figure 6a plots the two unordered bounds along with the tight bound when \( q = 0.99 \) (case (a) of proposition 4.1), where the tight bound is valid for all \( k \) in \([3,n]\), while figure 6b compares the bounds when \( q = 0.1 \) (case (b) of proposition 4.1) for \( k \geq 12 \) as the tight bound is valid when \( k \geq \lceil 2 + (n - 2)p/q \rceil = \lceil 11.8 \rceil = 12 \). The unordered Boros and Prekopa bound is much tighter than the unordered Schmidt, Siegel and Srinivasan bound in both figures. Hence, Figure 6 demonstrates that with ordering, the relative improvement of the Schmidt, Siegel and Srinivasan bound is much better than that of the Boros and Prekopa bound although both the ordered bounds reduce to the tight bound in (4.6).

5 Conclusion

In this paper we have provided results towards finding tight probability bounds for the sum of pairwise independent random variables. In Section 2, it was established through Theorem 2.1 that with pairwise independence, the Kounias [1968], Hunter [1976] and Worsley [1982] bound are tight for any \( p \in [0,1]^n \), which, to the best of our knowledge, has not been shown thus far in the literature dedicated to this topic. In fact, paraphrasing from Boros et al. [2014] (section 1.2), “As far as we know, in spite of the several studies dedicated to this problem, the complexity status of this problem, for feasible input, seems to be still open even for bivariate probabilities”. With pairwise independent random variables, feasibility is guaranteed and Theorem 2.1 shows that the tightest upper bound is computable in polynomial time (in fact in a simple closed form), thus providing a partial positive answer towards this question. In Section 3, for identical probabilities \( p \in [0,1] \) and any \( k \in [n] \), we used a constructive proof exploiting the symmetry in the problem, to identify the best upper bound \( P(n,k,p) \) in closed form and a corresponding extremal distribution. We further demonstrated the usefulness of this result by identifying instances when existing bounds are tight. In Section 4, when \( k \geq 2 \), we proposed new bounds exploiting ordering of the probabilities, which are at least as good as the unordered bounds and demonstrated instances when these ordered bounds are tight. To the best of our knowledge the idea of ordering has not been exploited thus far to tighten probability bounds for pairwise independent random variables.

We believe several interesting research questions arise from this work that need to be answered, two of which we list below:

(a) To the best of our knowledge, the computational complexity of evaluating (or approximating) the bound \( P(n,k,p) \) for general \( n, k \) and \( p \in [0,1]^n \) is still unresolved. While we provide the answer in closed form for \( k = 1 \), a natural question that arises is whether the tight bounds for general \( k \geq 2 \) with pairwise independent random variables are efficiently computable (or efficient to approximate)? We leave this for future research.

(b) The upper bound of 4/3 in Section 2.2 is derived for the ratio between the maximum probability for the union of arbitrarily dependent events and the probability of the union of pairwise independent events. We conjecture this upper bound is valid for the expected value of all non-decreasing, nonnegative submodular functions (of which the probability of the union is a special case) and leave it as an open question.
References


G. Boole. The Laws of Thought (1916 reprint), 1854.


