Ann Oper Res
https://doi.org/10.1007/s10479-018-2934-8

S.I.:STOCHASTIC MODELING AND OPTIMIZATION, IN MEMORY OF ANDRÁS PRÉKOPA

# Relaxations for probabilistically constrained stochastic programming problems: review and extensions

**Miguel A. Lejeune**[1] · **A. Prékopa**[2]



**Abstract** We consider probabilistically constrained stochastic programming problems, in which the random variables are in the right-hand sides of the stochastic inequalities defining the joint chance constraints. Problems of that kind arise in a variety of contexts, and are particularly difficult to solve for random variables with continuous joint distributions, because the calculation of the cumulative distribution function and its gradient values involves numerical integration and/or simulation in large dimensional spaces. We revisit known and provide new relaxations extensions to various probability bounding schemes that permit to approximate the feasible set of joint probabilistic constraints. The derived mathematical formulations relax the requirement to handle large multivariate cumulative distribution functions and involve instead the computation of marginal and bivariate cumulative distribution functions. We analyze the convexity of and computational challenges posed by the inferred relaxations

**Keywords** Stochastic programming · Probabilistic programming · Chance constraints · Relaxation and approximation problems · Reliability

## 1 Introduction

Let $x$ be the vector of decision variables and $\xi$ the vector of random variables. The following model

$$\min \; c^T x \tag{1a}$$
$$\text{s. to } \; Tx \geq \xi \tag{1b}$$

The paper was written in large part before Professor András Prékopa passed away.

✉ Miguel A. Lejeune
mlejeune@gwu.edu

1 Department of Decision Sciences, George Washington University, Washington, DC, USA

2 Rutgers University, Piscataway, NJ, USA

Published online: 18 June 2018

$$Ax \geq b \tag{1c}$$
$$x \geq 0 \tag{1d}$$

where $A$ is an $[m \times n]$-dimensional matrix, $T$ in an $[r \times n]$-dimensional matrix, $c$ and $x$ are $n$-dimensional vectors, $b$ and $\xi$ are $m$- and $r$-dimensional vectors, respectively, is the underlying program for static stochastic programming model formulations. One of them is referred to as programming under probabilistic constraints and is thereafter denoted by **SP1**

$$\begin{aligned} \textbf{SP1}: \ \min \ & c^T x \\ \text{s. to } & \mathbb{P}(Tx \geq \xi) \geq p \\ & (1c) - (1d), \end{aligned} \tag{2}$$

where $p \in (0, 1)$ is a prescribed probability level, usually close to 1. With $F$ the cumulative distribution function of $\xi$, the probabilistic constraint (2) can be equivalently rewritten as: $F(Tx) \geq p$.

The probabilistic constraint enforces the joint fulfillment of a system of linear inequalities with random right-hand side variables on or above a prescribed probability level $p$. Problems of that kind arise in multiple industrial and business contexts (see Prékopa 1995, 2003; Wallace and Ziemba 2005 and the references therein). Facing prediction uncertainty and having to design large scale planning strategies, it is very difficult for human decision makers to develop effective, coordinated solutions in real time. A possible consequence of this is the implementation of very conservative decisions leading to unnecessary delays and costs. In that respect, probabilistically constrained problems are very helpful to design ahead-of-time efficient strategies in the presence of uncertainty. Such problems are difficult to solve. If the random variables have continuous joint distributions, the calculation of the cumulative distribution function and of its gradient values is needed and is computationally intensive in general (Deák 1988, 2000; Genz 1992; Genz and Bretz 2009), since it requires numerical integration and/or simulation in high dimensional spaces (Guigues and Henrion 2017; Henrion and Möller 2012; Szántai 1988). The development of computationally tractable solution and reformulation methods is therefore critical.

Programming under probabilistic constraints, also called chance-constrained programming, was introduced by Charnes et al. (1958), who considered a set of $r$ individual probabilistic constraints (3) imposed on each particular stochastic inequality $T_i x \geq \xi_i$:

$$\begin{aligned} \textbf{SP2}: \ \min \ & c^T x \\ \text{s. to } & \mathbb{P}(T_i x \geq \xi_i) \geq p_i, \ i = 1, \ldots, r \\ & (1c)-(1d). \end{aligned} \tag{3}$$

The use of individual probabilistic constraints (3) is relatively easy to handle. However, in most situations, probabilistic constraints, taken individually, do not give an accurate representation of the considered system. Individual probabilistic constraints are only appropriate if the system is composed of components that do not affect each other. Miller and Wagner (1965) proposed a formulation for joint probabilistic constraints where the random variables $\xi_i, i = 1, \ldots, r$ are assumed to be independent. The probabilistic constraint (3) can then be rewritten as

$$\prod_{i=1}^{r} F_i(T_i x) \geq p, \tag{4}$$

where $F_i$ is the cumulative distribution function of $\xi_i$. The general case was introduced by Prékopa (1970, 1973, 1995). The problem in its most general form, referred to by **GSP**, can be stated as follows:

$$\begin{aligned} \textbf{GSP}: \ & \min \ c^T x \\ & \text{s. to } \mathbb{P}(g_i(x,\xi) \geq 0, \ i = 1,\ldots,r) \geq p \\ & \quad (1c) - (1d), \end{aligned} \tag{5}$$

where the stochastic inequalities $g_i(x,\xi) \geq 0, i = 1,\ldots,r$ are functions of $x$ and $\xi$, and $\xi$ is a random vector whose components are not necessarily independent. In the linear case where $g_i(x,\xi) = T_i x - \xi_i$, we obtain problem **SP1** whose chance constraint is not equivalent however to (3) and (4). In fact, the stochastic independence of $\xi$ is not assumed.

The first question to settle in connection with problem **GSP** is: under what conditions is the set of feasible solutions convex? A general theorem that answers this question is due to Prékopa (1973, 1995) and is stated below.

**Theorem 1** *If $g_i(x,\xi) \geq 0, i = 1,\ldots,r$ are concave functions of $x$ and $\xi$ and $\xi$ is a continuously distributed random vector with logarithmically concave (logconcave) probability density function, then*

$$\mathbb{P}(g_i(x,\xi) \geq 0, i = 1,\ldots,r)$$

*is a logarithmically concave function of $x$.*

Theorem 1 implies that if $\xi$ has continuous distribution with logarithmically concave probability density function, then the set of feasible solutions of problem **GSP** is convex. The collection of logconcave probability density functions is fairly extended and includes the multivariate non-degenerate normal distribution, the uniform (over a convex set) distribution, and, under some conditions on their parameters, the multivariate Dirichlet, Wishart, Gamma, Logistic, Exponential, Chi, Pareto distributions (see, e.g., Bagnoli and Bergstrom 2005; Mohtashami Borzadaran and Mohtashami Borzadaran 2011; Prékopa 1995, 2003). Note also that a logarithmically concave function is also quasi-concave and that the concavity assumption on the functions $g_i(x,\xi)$ can be replaced by the less restrictive and more general quasi-concavity requirement (see Prékopa 1995). For more convexity theorems concerning probabilistically constrained stochastic programming problems, the reader is referred to Prékopa (1995, 2003). There is no general result for discrete distributions that would parallel Prékopa's results (1973), except for the univariate case (Dentcheva et al. 2000; Prékopa 1995).

In this paper, we revisit and extend some bounding schemes that can be used to derive relaxations for the linear probabilistically constrained stochastic programming problem **SP1** in which the components of the random vector have some dependency structure and the stochastic inequalities are linear $g_i(x,\xi) = T_i x - \xi_i, i = 1,\ldots,r$.

We employ the term *relaxation*, since the proposed methods relax the requirement of computing large multivariate probabilities and their gradients and are based on the simpler calculation of marginal and bivariate probabilities associated with lower dimensional random vectors. Note however that the reformulations induced by the bounding schemes can be inner or outer approximations of **SP1**. We also study the computational challenges posed by the induced relaxations, study their convexity, and identify conditions under which the relaxations take the form of convex programming problems. The presentation is focused upon problems where the random variables follow continuous distributions although the proposed methods are also applicable to finitely distributed random variables. We do not review reformulation methods, such as the $p$-efficient concept (Dentcheva and Martinez 2013;

Dentcheva et al. 2000; Lejeune 2012b; Lejeune and Noyan 2010; Lejeune and Ruszczyński 2007; Prékopa 1990; Saxena et al. 2010), scenario-based reformulation with binary variables and knapsack constraints (Ruszczynski 2002), and pattern- and Boolean programming-based reformulations (Kogan and Lejeune 2013; Lejeune 2012a; Lejeune and Margot 2016), that are specific for discrete distributions. We refer the reader to Nemirovski and Shapiro (2006) for Bernstein-based approximations of chance constraints.

## 2 Relaxations based on bounding methods

The relaxation methods presented in this section cover five probability bounding schemes respectively based on the Boole-Bonferroni inequalities (Sect. 2.1), the binomial moment (Sect. 2.2), product-type inequalities (Sect. 2.3), the setwise dependence concept (Sect. 2.4), and the sum of disjoint products concept (Sect. 2.5).

### 2.1 Relaxations with Boole-Bonferroni bounding inequalities

#### *2.1.1 First-order relaxations with Boole's inequality*

We first review the Boole's approximation method for probabilistically constrained stochastic programming problems. Let $A_i$[1] be the event defined as $A_i : \xi_i \leq T_i x, i = 1, \ldots, r$ and $\overline{A}_i$ designate the complement of $A_i$. Combining Boole's inequality

$$\mathbb{P}\left(\bigcup_{i=1}^{r} A_i\right) \leq \sum_{i=1}^{r} \mathbb{P}(A_i) \tag{6}$$

and De Morgan's inequality

$$\bigcap_{i=1}^{r} A_i = \overline{\bigcup_{i=1}^{r} \overline{A}_i} \;\Leftrightarrow\; \mathbb{P}\left(\bigcap_{i=1}^{r} A_i\right) = 1 - \mathbb{P}\left(\bigcup_{i=1}^{r} \overline{A}_i\right) \tag{7}$$

that link marginal probabilities, and probabilities on the union and intersection of events, we obtain a lower bound on the probability of the intersection of multiple events

$$\mathbb{P}\left(\bigcap_{i=1}^{r} A_i\right) \;\geq\; 1 - \sum_{i=1}^{r} \mathbb{P}\left(\overline{A}_i\right) \;=\; \sum_{i=1}^{r} \mathbb{P}(A_i) - r + 1, \tag{8}$$

from which the following inner approximation of problem **SP1** can then be generated:

$$\begin{aligned} \textbf{BO1}: \;\; & \min \; c^T x \\ & \text{s. to} \; \sum_{i=1}^{r} F_i(T_i x) - r + 1 \geq p \\ & \qquad (1c)-(1d). \end{aligned} \tag{9}$$

A few comments are in order to compare the formulations **SP1** with joint chance constraint and its inner approximation **BO1**. First, while problem **SP1** requires the computation of multivariate probabilities, problem **BO1** alleviates the computational burden by computing

[1] The notation $A_i(x)$ may be used to underline the dependency on the decisions $x$. We drop the parenthesis $(x)$ to ease the notations as in Prékopa (2003).

marginal cumulative probabilities. Second, assuming that $\xi$ has a quasi-concave probability distribution, it follows from Theorem 1 that **SP1** is a convex programming problem. However, Theorem 1 does not apply directly to **BO1** since the summation operation does not necessarily preserve quasi-concavity. Thus, even if each function $F_i(T_i x)$, $i = 1, \ldots, r$ is quasi-concave (or log-concave), there is no guarantee that their summation $\sum_{i=1}^{r} F_i(T_i x)$ is quasi-concave (or log-concave) and the feasible set defined (9) may thus not be convex.

There exists however some probability distributions for which problem **BO1** can be proven convex, as indicated by Prékopa (2001). Let $\mu_i$, $i = 1, \ldots, r$ be the mean of $\xi_i$, $F_i^{-1}$ be the inverse of the marginal probability distribution of $\xi_i$, and $F_i^{-1}(p)$ be the $p$-quantile of $F_i$. We have the following result for the normal distribution.

**Lemma 1** *Let $p \in [0.5, 0)$. If the $r$-variate random variable $\xi$ is normally distributed, problem* **BO1***, in which we introduce the implied constraints*

$$F_i(T_i x) \geq p \ , \ i = 1, \ldots, r \ , \tag{10}$$

*is a convex programming problem.*

Und der Grund, warum (10) aus (9) folgt, ist folgender: Die Ungleichung in (9) sagt, dass die Summe von den r vielen F_i-Termen mindestens r - 1 + p sein muss. Wenn jetzt einer der F_i-Terme kleiner als p wäre, dann wäre es unmöglich, r - 1 + p zu erreichen, weil alle anderen F_i-Terme jeweils höchstens 1 und damit in Summe höchstens r - 1 sind. Also muss jeder einzelne F_i-Term mindestens p sein.

*Proof* Note first that (9) implies (10), whose insertion does not alter the feasible set of **BO1**.

For $p \leq 0.5$, the first derivative of the standard normal density function is negative, therefore implying that the cumulative function of the standard normal distribution is concave and that the feasible set of each constraint (10) is convex, since it requires a concave function to be at least equal to a constant.

Furthermore, since the sum of concave functions is concave, it follows that $\sum_{i=1}^{r} F_i(T_i x)$ is concave and that the feasible set defined by (9) is convex too, which provides the result that we set out to prove. □

Note that similar results are not exclusive for the normal distribution and can be obtained for other continuous distributions, such as the Dirichlet distribution [see results derived by (Prékopa 2001)].

We can derive a second inner approximation formulation (Prékopa 1995, 1999), equivalent to **BO1**, by lifting the decisional and constraint spaces. Let $q_1, \ldots, q_r$ be auxiliary non-negative decision variables that will be used to rewrite the nonlinear constraint (9).

$$\textbf{BO2}: \ \min \ c^T x$$

$$\text{s. to} \ F_i(T_i x) \geq q_i \ , \ i = 1, \ldots, r \tag{11a}$$

$$1 - p \geq \sum_{i=1}^{r} (1 - q_i) \tag{11b}$$

$$p \leq q_i \leq 1 \ , \ i = 1, \ldots, r$$

$$(1c)-(1d). \tag{11c}$$

The formulation **BO2** includes $r$ individual chance constraints (11a). The constraints (11c) are implied by (11b) and can be removed. They can be helpful to speed up the solution of problem **BO2**. While $p$ is a fixed probability level, note that $q_i$, $i = 1, \ldots, r$ are decision variables. Problem **BO2** is also convex under the normality condition and value of the probability level specified in Lemma 1.

A third inner approximation can be derived (see, e.g., Chen et al. 2010). It takes the form of the linear programming problem

$$
\begin{aligned}
\textbf{BO3}: \ & \min \ c^T x \\
& \text{s. to } T_i x \geq F_i^{-1}(1-(1-p)/r) \ , \ i=1,\dots,r \\
& \quad (1c)-(1d).
\end{aligned} \tag{12}
$$

While problem **BO3** is easier to solve than its nonlinear counterparts **BO1** and **BO2**, **BO3** is (much) more conservative than **BO1** and **BO2** as shown in Lemma 2, in particular when the size $r$ of the random vector increases.

**Lemma 2** *Problem* **BO3** *is an inner approximation of* **BO1** *and* **BO2**. *The feasible set of* **BO1** *includes the feasible set of* **BO3***:*

$$
H_3 = \left\{x \in \mathbb{R}^n_+ : (12)\right\} \ \subseteq \ H_1 = \left\{x \in \mathbb{R}^n_+ : (9)\right\}. \tag{13}
$$

*Proof* Consider the feasible solution of **BO3** in which all inequalities (12) are binding: $F_i(T_i x) = 1-(1-p)/r, i=1,\dots,r$. It is easy to see that this solution, and hence all feasible solutions for **BO3**, satisfy (9) and are feasible for **BO1**. Therefore, $H_3 \subseteq H_1$. Similarly, any feasible solution of **BO3** can be mapped into a solution feasible for **BO2**. □

Note that the converse $H_1 \subseteq H_3$ is in general not true. Consider a solution $\bar{x}$ feasible for **BO1** such that $F_i(T_i\bar{x}) = 1-(1-p)/r, i \in \{1,\dots,r\}\backslash\{i', i''\}$, $F_{i'}(T_{i'}\bar{x}) = 1-(1-p)/r+s$, $F_{i''}(T_{i''}\bar{x}) = 1-(1-p)/r-s$, $s>0$, and $1-(1-p)/r-s \geq p$. Obviously, $\bar{x} \in H_1$ and $\bar{x} \notin H_3$, which highlights the stated result.

### 2.1.2 Second-order bounding problems with Boole-Bonferroni's inequalities

An appealing feature of the first-order bounds based on Boole's inequality is that they do not require the knowledge of the covariance structure. However, they can be conservative, in particular under strong positive dependence between the components of the random vector. In this section, we propose relaxation methods using several variants of the second-order Boole-Bonferroni inequalities on the probability of the union of a finite number of events.

The basic second degree Boole-Bonferroni lower bound on the probability of the union of a number of events is the left term in (14):

$$
\sum_{i=1}^{r} \mathbb{P}(A_i) - \sum_{i=1}^{r-1} \sum_{j=i+1}^{r} \mathbb{P}(A_i, A_j) \leq \mathbb{P}\left(\bigcup_{i=1}^{r} A_i\right) \leq \sum_{i=1}^{r} \mathbb{P}(A_i), \tag{14}
$$

for which Kwerel (1975) proposed a tighter variant:

$$
\mathbb{P}\left(\bigcup_{i=1}^{r} A_i\right) \leq \sum_{i=1}^{r} \mathbb{P}(A_i) - \frac{2}{r} \sum_{i=1}^{r-1} \sum_{j=i+1}^{r} \mathbb{P}(A_i, A_j). \tag{15}
$$

Subsequent improvements were obtained using graph theory. The upper bound (16) on the probability of the union of a finite number of events presented in Theorem 2 (Worsley 1982) is obtained by removing, from the sum of the $r$ (event) marginal probabilities, $(r-1)$ appropriately chosen bivariate probabilities of pairs of events.

**Theorem 2** *Worsley (1982) Let the events $A_i, i=1,\dots,r$ be the vertices of a graph $G$, where two vertices $A_i$ and $A_j$ are linked by an edge $e_{ij}$ if and only if they are not mutually exclusive. If $\mathcal{T}$ is a subgraph of $G$, then*

$$
\mathbb{P}\left(\bigcup_{i=1}^{r} A_i\right) \leq \sum_{i=1}^{r} \mathbb{P}(A_i) - \sum_{\{(i,j): e_{ij} \in \mathcal{T}\}} \mathbb{P}(A_i, A_j). \tag{16}
$$

*is valid if and only if $\mathcal{T}$ is a tree.*

Hunter (1976) showed that the right-hand side of (16) is minimal and the upper bound is the sharpest if we construct the maximum or heaviest spanning tree $\mathcal{T}^*$ in which each edge $e_{ij}$ is weighted by the bivariate probability $\mathbb{P}(A_i, A_j)$ of the two events it connects:

$$\mathcal{T}^* = \left\{ (i, j) \in \mathcal{T} : \max \sum_{(i,j)\in\mathcal{T}} \mathbb{P}(A_i, A_j) : \mathcal{T} \text{ is a spanning tree} \right\}. \tag{17}$$

Hunter's approach has been used to derive upper bounds on the probability of the union and to obtain approximate solutions for probabilistically constrained programming problems (see Prékopa 2003). Within this approach, the selection of the edges $e_{ij}$ included in the maximal spanning tree $\mathcal{T}^*$ depends on the bivariate probabilities $\mathbb{P}(A_i, A_j) = \mathbb{P}(\xi_i \leq T_i x, \xi_j \leq T_j x)$, which themselves are not known prior to the solution of the optimization problem incorporating Hunter's bound. The values of their bivariate probabilities and the inclusion or not of the corresponding edges in the heaviest spanning tree can be viewed as endogeneous or decision-dependent unknowns as they depend on the optimal solution and the values assigned to the decision variables $x$. Thus, the weighting of the edges by the probabilities does not allow the a priori construction of the maximum spanning tree and Hunter's bound cannot be conveniently included in the solution process. By contrast, the following inequality (Worsley 1982)

$$\mathbb{P}\left(\bigcup_{i=1}^{r} A_i\right) \leq \sum_{i=1}^{r} \mathbb{P}(A_i) - \sum_{i=1}^{r-1} \mathbb{P}(A_i, A_{i+1}). \tag{18}$$

is not optimal in the sense that it does not guarantee the provision of the sharpest bound, but has the advantage that it can be used for the a priori determination of the probabilities used to obtain an upper bound for $\mathbb{P}(\bigcup_{i=1}^{r} A_i)$ and, subsequently, for the solution of the inner approximation of problem **SP1**.

As compared to (15), Worsley's inequality (18) is more practical, since it contains a smaller number of bivariate probability terms than (15). The difference between the number of bivariate probability terms included in (15) and (18) increases monotonically with the size $r$ of the random vector.

We shall now propose a new procedure to derive a heaviest spanning tree $\tilde{\mathcal{T}}$ which approximates the optimal spanning tree $\mathcal{T}^*$ (17) and is based on the correlation levels $\rho_{ij}$ between two events $A_i$ and $A_j$.

**Lemma 3** *Let $\tilde{\mathcal{T}}$ be the maximal spanning tree defined by:*

$$\tilde{\mathcal{T}} = \left\{ (i, j) \in \mathcal{T} : \max \sum_{(i,j)\in\mathcal{T}} \rho_{ij} : \mathcal{T} \textit{ is a spanning tree} \right\}. \tag{19}$$

*The inequality*

$$\mathbb{P}\left(\bigcup_{i=1}^{r} A_i\right) \leq \sum_{i=1}^{r} \mathbb{P}(A_i) - \sum_{\{(i,j):e_{ij}\in\tilde{\mathcal{T}}\}} \mathbb{P}(A_i, A_j). \tag{20}$$

*is valid and provides an upper bound on the probability of the union of $r$ events.*

The proof follows from Theorem 2. because it is a spanning tree

The sharpest bound (17) of type (16) is obtained by using Hunter's procedure which involves the maximization of the bivariate probabilities (see (17)). We opted to maximize the sum of the correlations, since it represents (part of) the information captured by the sum of the bivariate probabilities. We note that criteria other than the maximization of the sum of the correlations (see (19)) can be considered to build a maximal spanning tree.

In contrast to bivariate probabilities used in the spanning tree underlying Hunter's bound, correlation levels are fixed and do not depend on the decision variables $x$. Therefore, the spanning tree $\tilde{\mathcal{T}}$ and the selection of the edges it includes can be determined prior to the formulation and solution of the relaxation problem incorporating Hunter's bound. Let $\overline{A}_i$ be the complement of $A_i$: $\mathbb{P}(\overline{A}_i) = \mathbb{P}(\xi_i > T_i x)$ and $F_{ij}$ denote the bivariate cumulative distribution function of $\xi_i$ and $\xi_j$.

**Theorem 3** *The inequality*

$$\sum_{i=1}^{r} F_i(T_i x) - \sum_{\{(i,j):e_{ij}\in\tilde{\mathcal{T}}\}} \left(F_i(T_i x) + F_j(T_j x) - F_{ij}(T_i x, T_j x)\right) \geq p \tag{21}$$

*implies that* (2) *holds. The left-hand side of* (21) *limits from below the probability on the intersection of $r$ events.*

*Proof* We have successively that:

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{i=1}^{r} A_i\right) &= 1 - \mathbb{P}\left(\overline{A}_1 \cup \ldots \cup \overline{A}_r\right) && (22)\\
&\geq 1 - \left(\sum_{i=1}^{r} \mathbb{P}(\overline{A}_i) - \sum_{\{(i,j):e_{ij}\in\tilde{\mathcal{T}}\}} \mathbb{P}(\overline{A}_i, \overline{A}_j)\right) && (23)\\
&= 1 - \left(\sum_{i=1}^{r}(1 - \mathbb{P}(A_i)) - \sum_{\{(i,j):e_{ij}\in\tilde{\mathcal{T}}\}} \left(1 - \mathbb{P}(A_i \cup A_j)\right)\right) && (24)\\
&= 1 - r + \sum_{i=1}^{r} \mathbb{P}(A_i) + (r-1) - \sum_{\{(i,j):e_{ij}\in\tilde{\mathcal{T}}\}} \mathbb{P}(A_i \cup A_j) && (25)\\
&= \sum_{i=1}^{r} \mathbb{P}(A_i) - \sum_{\{(i,j):e_{ij}\in\tilde{\mathcal{T}}\}} \left(\mathbb{P}(A_i) + \mathbb{P}(A_j) - \mathbb{P}(A_i \cap A_j)\right), && (26)
\end{aligned}$$

where (22) is due to De Morgan's inequality, (23) follows from (16), (24) is based on the complement rule, (25) is valid since the cardinality of $\tilde{\mathcal{T}}$ is $r-1$, and (26) is equivalent to the left-hand side of (21). This implies that the left-hand side of (21) is a lower bound for $\mathbb{P}(\bigcap_{i=1}^{r} A_i)$, which validates (21). □

Corollary 1 follows immediately.

**Corollary 1** *The nonlinear optimization problem*

$$\begin{aligned}
&\min\ c^T x\\
&\textit{s. to}\ \ (1c)\text{–}(1d);\ \ (21)
\end{aligned}$$

*is an inner approximation of* **SP1** *and its optimal value is an upper bound on the optimal value of* **SP1**.

The above problem is in general non-convex. There exists some cases for which the above problem can be proven convex using Theorem 4 (Prékopa 2001). Let $R$ denote the nonsingular correlation matrix of a multivariate random variable.

**Theorem 4** *Prékopa (2001) The $r$-variate standard normal probability distribution function $\Phi(z; R)$ is concave in the set:*

like mu, but for corr matrix

$$\left\{z : z_i \geq \sqrt{r-1},\ i = 1, \ldots, r\right\}. \tag{27}$$

We can now derive a second lower relaxation for **SP1** and determine a set of conditions under which it is convex. Let $\mu_i$ and $\sigma_i$ be the mean and standard deviation of $\xi_i$, $i = 1, \ldots, r$.

**Theorem 5** *Let $\xi$ be a $r$-variate normally distributed random variable and $p \in [0.841345, 1)$. Problem* **BH1**

$$\begin{aligned} \mathbf{BH1}:\ & \min\ c^T x \\ & \text{s. to} \sum_{\{(i,j):e_{ij} \in \tilde{\mathcal{T}}\}} F_{ij}(T_i x, T_j x) \geq p + r - 2 \\ & (1c) - (1d) \end{aligned} \tag{28}$$

is a convex inner approximation of **SP1** and provides an upper bound on the optimal value of **SP1**.

*Proof* We want to demonstrate that (28) is an inner approximation of (21) and derive an upper bound for

$$\sum_{\{(i,j):e_{ij} \in \tilde{\mathcal{T}}\}} \left(F_i(T_i x) + F_j(T_j x)\right) - \sum_{i=1}^{r} F_i(T_i x)\,. \tag{29}$$

Since $\tilde{\mathcal{T}}$ is a spanning tree, we have that: 1) $\tilde{\mathcal{T}}$ includes exactly $(r-1)$ edges; hence $|\tilde{\mathcal{T}}| = r-1$ and $\sum_{\{(i,j):e_{ij} \in \tilde{\mathcal{T}}\}} \left(\mathbb{P}(A_i) + \mathbb{P}(A_j)\right) \leq 2(r-1)$; and 2) each event $A_i$ appears at least once in the summation $\sum_{\{(i,j):e_{ij} \in \tilde{\mathcal{T}}\}} \left(\mathbb{P}(A_i) + \mathbb{P}(A_j)\right)$. Combining 1) and 2), it follows that after substracting $\sum_{i=1}^{r} F_i(T_i x)$ from $\sum_{\{(i,j):e_{ij} \in \tilde{\mathcal{T}}\}} \left(F_i(T_i x) + F_j(T_j x)\right)$, the number of terms in (29) is equal to $r-2$, each equal to 1 at most, which implies that

$$\sum_{\{(i,j):e_{ij} \in \tilde{\mathcal{T}}\}} \left(\mathbb{P}(A_i) + \mathbb{P}(A_j)\right) - \sum_{i=1}^{r} \mathbb{P}(A_i) \leq r - 2 \qquad \text{(29a)}$$

always holds. Therefore, (28) implies (21) and the above problem is an inner approximation of **SP1**.

Next, we observe that each function $\sum_{\{(i,j):e_{ij} \in \tilde{\mathcal{T}}\}} F_{ij}(T_i x, T_j x)$ in (28) is concave on the set $\left\{x : \frac{T_i x - \mu_i}{\sigma_i} \geq 1, \frac{T_j x - \mu_j}{\sigma_j} \geq 1\right\}$. This is the case if $F(T_i x) \geq 0.841345$ and $F(T_j x) \geq 0.841345$, which is guaranteed since $p \in [0.841345, 1)$. Therefore, since the sum of concave functions is concave, (28) has a convex feasible region. ◻

### 2.2 Binomial moment bounding scheme

The first formulation based on the binomial moment problem was proposed by Prékopa (1999) and was formulated for a finite number of events $A_1, \ldots, A_r$ defined in a specified probability space. It permits to derive sharp lower and upper bounds for the probability of the following Boolean functions of events: $\bigcup_{i=1}^{r} A_i$ and $\bigcap_{i=1}^{r} A_i$.

Denoting by $\nu$ the number of events that occur out of $A_1, \ldots, A_r$, it is well known that

binomial coefficient - num of ways to choose v elements from a set of k elements = n! divided by j!(n-j)!

$$\mathbb{E}\left[\binom{\nu}{k}\right] = S_k \ , k = 1, \ldots, r, \tag{30}$$

where $S_k$ represents the $k^{th}$ binomial moment of the random variable defined as:

$$S_k = \sum_{1 \le i_1 < \ldots < i_k \le r} \mathbb{P}\left(A_{i_1} \cap \ldots \cap A_{i_k}\right), \ k = 1, \ldots, r. \tag{31}$$

S2, 4 events: we choose possible 2-combinations such that 1<=i1<=i2<=4 - (1,2), (1,3), (1,4), (2,3), (2,4), (3,4)
s2 = P(A1 disj A2) + P(A1 disj A3)...

In particular, $S_1 = \sum_{i=1}^{r} \mathbb{P}(A_i)$ is the sum of the marginal probabilities.

Let $v_i = \mathbb{P}(\nu = i), i = 0, 1, \ldots, r$, we can reformulate (30) as

$$\sum_{i=0}^{r} \binom{i}{k} v_i = S_k, k = 0, \ldots, r, \tag{32}$$

in which the values of $v_0, \ldots, v_r$ and $S_1, \ldots, S_r$ uniquely determine each other, and $S_0 = 1$.

Consider that only $S_1, \ldots, S_m$ are known or can be computed conveniently and let $m < r$ be the maximum number of (event) intersections considered. The binomial moment problem is defined as the following linear programming problem (Prékopa 2003):

$$\min(\max) \ \sum_{i=0}^{r} f_i v_i \tag{33a}$$

$$\text{s. to} \ \sum_{i=0}^{r} \binom{i}{k} v_i = S_k \ , \ k = 0, \ldots, m \tag{33b}$$

$$v_i \ge 0 \ , \ i = 0, \ldots, r \tag{33c}$$

where $f_0, f_1, \ldots, f_m$ are constants.
If

$$f_i = \begin{cases} 0 \text{ if } i = 0 \\ 1 \text{ if } i > 0 \end{cases}, \tag{34}$$

then the optimal value of the minimization (resp., maximization) problem (33) provides a lower (resp., upper) bound for the probability $\mathbb{P}(\bigcup_{i=1}^{r} A_i)$. If

$$f_i = \begin{cases} 0 \text{ if } i < r \\ 1 \text{ if } i = r \end{cases}, \tag{35}$$

then the optimal value of the minimization (resp., maximization) problem (33) provides a lower (resp., upper) bound for the joint probability $\mathbb{P}(\bigcap_{i=1}^{r} A_i)$. Given that $\mathbb{P}(\bigcap_{i=1}^{r} A_i) = 1 - \mathbb{P}(\bigcup_{i=1}^{r} \overline{A}_i)$, the lower (resp., upper) bound for $\mathbb{P}(\bigcap_{i=1}^{r} A_i)$ obtained with the minimization (resp., maximization) problem (33) and the parametrization (35) is equal to 1- the upper (res., lower) bound for $\mathbb{P}(\bigcup_{i=1}^{r} \overline{A}_i)$ obtained with the maximization (resp., minimization) problem (33) in which each $S_k$ is replaced by its complement $\overline{S}_k$ and the parametrization (34).

We shall now use the bounds on $\mathbb{P}(\bigcap_{i=1}^{r} A_i)$ to relax problem **SP1**. Let $F_{i_1,\ldots,i_k}$ designate the joint probability distribution function of the random variables $\xi_{i_1},\ldots,\xi_{i_k}$ and $\alpha$ be an arbitrary positive number.

Prékopa (1999) has shown that the two following problems

$$\textbf{BI1}: \ \min \ c^T x + \alpha v_r \tag{36a}$$

$$\text{s. to} \ \sum_{i=0}^{r} v_i = 1 \tag{36b}$$

$$v_1 + 2v_2 + \ldots + r v_r = \sum_{i=1}^{r} F_i(T_i x) \tag{36c}$$

$$v_2 + \binom{3}{2} v_3 + \ldots + \binom{r}{2} v_r = \sum_{1\le i_1<i_2\le r} F_{i_1,i_2}(T_{i_1}x, T_{i_2}x) \tag{36d}$$

$$\ldots$$

$$v_m + \binom{m+1}{m} v_{m+1} + \cdots + \binom{r}{m} v_r = \sum_{1\le i_1<\cdots<i_m\le r} F_{i_1,\ldots,i_m}(T_{i_1}x,\ldots,T_{i_m}x) \tag{36e}$$

$$v_r \ge p$$

$$(1c)-(1d); \ (33c), \tag{36f}$$

and

$$\textbf{BI2}: \ \max \ -c^T x + \alpha v_r$$

$$\text{s. to} \ (1c)-(1d); \ (33c); \ (36b)-(36f). \tag{37a}$$

are outer approximations of problem **SP1**.

The above problems are in general non-convex since, first, we have nonlinear equality constraints and second the right-hand sides of the nonlinear constraints involve the sum of probability distribution functions. Many multivariate probability distributions have quasi-concave or log-concave density functions, which implies that their marginal distributions are also quasi-concave or log-concave. However, the summation is not an operation preserving the quasi-concavity and log-concavity properties and the sums of quasi-concave or log-concave functions is not necessarily quasi-concave or log-concave, which impacts the statements that can be made about the convexity of their feasible area.

Next, we restrict our attention to the first two binomial moments $S_1$ and $S_2$ and derive a convex programming relaxation of problem **BI1**. Let $L$ be an infinitesimal number.
close to 0, but not 0

**Theorem 6** *If $\xi$ has a nondegenerate $r$-variate normal distribution and $p \ge [0.841345, 1)$, problem* **BI3**

$$\textbf{BI3}: \ \min \ c^T x$$

$$s. \ to \ \sum_{i=1}^{r} F_i(T_i x) + \epsilon_1 \ge \sum_{i=0}^{r} i v_i \tag{38a}$$

$$\sum_{1\le i_1<i_2\le r} F_{i_1,i_2}(T_{i_1}x, T_{i_2}x) + \epsilon_2 \ge \sum_{i=2}^{r} \binom{i}{2} v_i \tag{38b}$$

$$F_i(T_i x) \ge p \qquad i = 1,\ldots,r \tag{38c}$$

$$F_{i_1,i_2}(T_{i_1}x, T_{i_2}x) \geq p \qquad 1 \leq i_1 < i_2 \leq r \qquad (38d)$$

$$-L \leq \epsilon_k \leq L \qquad k = 1, 2 \qquad (38e)$$

$$(1c)–(1d);\ (33c);\ (36b);\ (36f) \qquad .$$

*is a convex programming relaxation of* **BO1** *and its optimal value* $(\mathbf{x}, \mathbf{v})$ *provides a lower bound* $c^T\mathbf{x}$ *on the optimal value of* **SP1**.

*Proof* In the linear case $g_i(x, \xi) = T_i x - \xi_i, i = 1, \ldots, r$, (5) reads $\mathbb{P}(T_i x \geq \xi_i, i = 1, \ldots, r) = F(T_i x, \ldots, T_r x) \geq p$ and it is necessary (but not sufficient) for it to hold to have $F_i(T_i x) \geq p, i = 1, \ldots, r$ and $F_i(T_{i_1}x, T_{i_2}x) \geq p, i \leq i_1 < i_2 \leq r$. Therefore, (38c) and (38d) are valid and their incorporation does not shrink the feasible set of **SP1**.

Theorem 4 indicates that each function $F_i(T_i x)$ in the left-hand side of (38a) and (38c) is concave on the set $\{x : T_i x - \mu_i \geq 0\}$. This is equivalent to requiring that $T_i x \geq 0 \Leftrightarrow F(T_i x) \geq 0.5$, which is ensured by (38c) given that $p \geq 0.841345$. Therefore, each function $F_i(T_i x)$ is concave. Since the sum of concave functions is concave, the left-hand-side of (38a) is a concave function and the feasible set defined by (38a) is convex.

Similarly, it follows from Theorem 4 that each function $F_{i_1,i_2}(T_{i_1}x, T_{i_2}x)$ in (38b) and (38d) is concave on the set $\left\{x : \frac{T_{i_1}x-\mu_{i_1}}{\sigma_{i_1}} \geq 1, \frac{T_{i_2}x-\mu_{i_2}}{\sigma_{i_2}} \geq 1\right\}$. This is the case if $F(T_{i_1}x) \geq 0.841345$ and $F(T_{i_2}x) \geq 0.841345$, which is ensured by (38c) given that $p \geq 0.841345$. Therefore, each function $F_{i_1,i_2}(T_{i_1}x, T_{i_2}x)$ is concave. Given that the sum of concave functions is concave, (38b) has a convex feasible region.

Since the intersection of convex sets is convex, problem **BI3** is a convex programming problem. □

Finally, note that Theorem 6 remains valid under more encompassing conditions than those enunciated in its statement. First, Prékopa (2001) notes that concavity results similar to those presented in Theorem 4 for the normal distribution could be obtained for other multivariate continuous distributions (*e.g.*, Dirichlet distribution). Second, the lower bound $\sqrt{r-1}$ in (27) is not always the sharpest one. For example, in the case of a bivariate standard normal distribution with independent marginals, the concavity of the probability distribution function is ensured if $z_i \geq 0.51, i = 1, 2$ (Prékopa 2001), while Theorem 4 requires $z_i \geq 1, i = 1, 2$. This means that the condition $p \geq 0.841345$ in Theorem 6 is a conservative one and that the applicability of Theorem 6 could be extended to lower values of $p$. Note that constraints (38e) permit to relax constraints (36c) and (36d) and to limit their violation to an infinitesimal value $\epsilon_1$ or $\epsilon_2$.

### 2.3 First-order bounding scheme with Slepian's product-type inequality

Slepian's inequality (Slepian 1962) is a first-order product-type inequality and was first stated as follows. If $R$ and $R'$ are two correlation matrices with the inequality $R \geq R'$ holding componentwise, then, for any $y = (y_1, \ldots, y_r)$ , we have

$$F(y; R) \geq F(y; R'), \tag{39}$$

where $F$ is the $r$-variate standard normal probability distribution function.
The $r$-variate random vector $\xi$ with $\mathbb{E}[\xi_i] = \mu_i, Var(\xi_i) = \sigma_i^2, i = 1, \ldots, r$, and correlation matrix $R$ can be standardized as follows:

$$F\left(\frac{T_i x - \mu_i}{\sigma_i},\ i = 1, \ldots, r; R\right). \tag{40}$$

If $R \geq R'$, then we have the inequality

$$F\left(\frac{T_i x - \mu_i}{\sigma_i},\ i = 1, \ldots, r; R\right) \geq F\left(\frac{T_i x - \mu_i}{\sigma_i},\ i = 1, \ldots, r; R'\right). \tag{41}$$

If we impose a probabilistic requirement, via a chance constraint, on the right-hand side term in (41), rather than on the left-hand side term, and substitute it for the probabilistic constraint in (2), then the set of feasible solutions becomes smaller and the optimal value larger. Clearly, the optimal value of

$$\begin{aligned}\mathbf{S1}:\ &\min\ c^T x\\ &\text{s. to}\ F\left(\frac{T_i x - \mu_i}{\sigma_i},\ i = 1, \ldots, r; R'\right) \geq p\\ &(1c){-}(1d)\end{aligned} \tag{42}$$

provides us with an upper bound on the optimal value of problem **SP1**.

Slepian's inequality is especially useful if the correlations between the different random variables are all non-negative or non-positive, because the use of independent random variables can then provide sharper bounds. Consider, for example, that $R \geq 0$ and set $R'$ as the identity matrix $R' = I$ (*i.e.*, which implies $R \geq R'$). Then, Slepian's inequality (41) can be successively rewritten as:

$$\begin{aligned}&F\left(\frac{T_i x - \mu_i}{\sigma_i},\ i = 1, \ldots, r; R\right) \geq F\left(\frac{T_i x - \mu_i}{\sigma_i},\ i = 1, \ldots, r; R'\right)\\ &= \prod_{i=1}^{r} F_i\left(\frac{T_i x - \mu_i}{\sigma_i}\right) &(43)\\ &\Leftrightarrow ln\left(F\left(\frac{T_i x - \mu_i}{\sigma_i},\ i = 1, \ldots, r; R\right)\right) \geq \sum_{i=1}^{r} ln\left(F_i\left(\frac{T_i x - \mu_i}{\sigma_i}\right)\right). &(44)\end{aligned}$$

**Theorem 7** *If the random variable $\xi$ has a lognormal density function with correlation matrix $R \geq 0$, problem* **S2**

$$\begin{aligned}\mathbf{S2}:\ &\min\ c^T x\\ &s.\ to\ \sum_{i=1}^{r} ln\left(F_i\left(\frac{T_i x - \mu_i}{\sigma_i}\right)\right) \geq ln(p)\\ &(1c){-}(1d)\end{aligned} \tag{45}$$

*is a convex inner approximation of Problem* **S1** *and its optimal value provides an upper bound on that of* **SP1**.

*Proof* The inner approximation and upper bound statements follow from (44) and (45). The convexity results from Theorem 1. Since $\xi$ has a logconcave density function, each term in the left-hand side of (45) is a concave function. The concavity property carries over the summation operation, which implies that the left-hand side of (45) is a concave function and that (45) defines a convex feasible region. □

We can use the optimal solution $\mathbf{x^I}$ of **S2** to derive a valid inequality for **SP1**.

Let $\mathbf{w^I} = F\left(\frac{T_i \mathbf{x^I} - \mu_\mathbf{i}}{\sigma_i},\ i = 1, \ldots, r;\ R\right)$. The inequality

$$\mathbf{w^I} \geq F\left(\frac{T_i x - \mu_i}{\sigma_i},\ i = 1, \ldots, r;\ R\right)$$

is therefore valid for **SP1**. However, its feasible area is not convex.

### 2.4 Setwise bounding schemes dependence bounding problems

We shall now introduce a generalization of Slepian's inequality to derive lower and upper bounds on the probability of the intersection of $r$ events. The bounds rely on the concept of set dependence (Chhetry et al. 1989; Costigan 1996). As Slepians's product-type inequality, setwise bounds require the verification of the positive dependence conditions. However these conditions are weaker than those needed to apply Slepian's inequality. Compared to the Boole-Bonferroni's inequality, they have the advantage of being non-degenerate and being exact under independence (Costigan 1996).

Define the $r$-variate random variable $\xi$ as the concatenation of a number $\upsilon$ of $r_k$-dimensional disjoint subvectors $\tilde{\xi}_k, k = 1, \ldots, \upsilon$:

$$\xi = \left(\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_\upsilon\right),$$

with $r = \sum_{k=1}^{\upsilon} r_k$ and such that

$$\mathbb{P}(\xi_i \leq T_i x, i = 1, \ldots, r) = \mathbb{P}(\tilde{\xi}_k \in L_k, k = 1, \ldots, \upsilon),$$

where each set $L_k, k = 1, \ldots, \upsilon$ is a collection of disjoint events.

Theorem 8 is due to Chhetry et al. (1989).

**Theorem 8** *Let $B_k = \{i : \xi_i \in \tilde{\xi}_k, i = 1, \ldots, r\}, k = 1, \ldots, \upsilon$ and $\rho_{ij}$ be the correlation between $\xi_i$ and $\xi_j$.*

*If $\tilde{\xi} = (\tilde{\xi}_1, \ldots, \tilde{\xi}_\upsilon)$ is normally distributed, then*

$$\mathbb{P}(\xi_i \leq T_i x, i = 1, \ldots, r) \geq \prod_{k=1}^{\upsilon} \mathbb{P}\left(\tilde{\xi}_k \in L_k\right) \tag{46}$$

*if*

$$\rho_{ij} \geq 0 \textit{ for each } i \in B_k \textit{ and } j \in B_{k'}, k \neq k'. \tag{47}$$

It follows that $\prod_{k=1}^{\upsilon} \mathbb{P}\left(\tilde{\xi}_k \in L_k\right)$ is a valid lower bound for $\mathbb{P}(\xi_i \leq T_i x, i = 1, \ldots, r)$. The condition (47) in Theorem 8 shows that the concept of setwise dependence imposes less stringent requirements on the positive (resp., negative) dependence between components of the random vectors than the Slepian's inequality. The above result holds if the subvectors $\tilde{\xi}_k, k = 1, \ldots, \upsilon$ are setwise positive lower orthant dependent (Costigan 1996). Corollary 2 and 3 follow immediately.

**Corollary 2** *Let $\tilde{\xi}_1, \ldots, \tilde{\xi}_\upsilon$ be such that (47) holds. Problem*

$$\begin{aligned} &\min\ c^T x \\ &\textit{s. to}\ \prod_{k=1}^{\upsilon} \mathbb{P}\left(\tilde{\xi}_k \in L_k\right) \geq p \\ &(1c){-}(1d) \end{aligned} \tag{48}$$

*is a nonlinear inner approximation of* **SP1** *and its optimal value is an upper bound on that of* **SP1**.

*Using the same argument, if* $\tilde{\xi}_1, \ldots, \tilde{\xi}_\upsilon$ *are such that*

$$\rho_{ij} \leq 0 \text{ for each } i \in B_k \text{ and } j \in B_{k'}, k \neq k', \tag{49}$$

*then*

$$\mathbb{P}(\xi_i \leq T_i x, i = 1, \ldots, r) \leq \prod_{k=1}^{\upsilon} \mathbb{P}\left(\tilde{\xi}_k \in L_k\right) \tag{50}$$

*and problem*

$$\begin{aligned} &\min\ c^T x \\ &\textit{s. to}\ \ (1c)-(1d);\ \ (48) \end{aligned}$$

*is a nonlinear relaxation of* **SP1** *and its optimal value is a lower bound on the optimal value of* **SP1**.

Consider a 4-variate random variable with correlation terms $\rho_{12}, \rho_{34} < 0$ and $\rho_{13}, \rho_{14}, \rho_{23}, \rho_{24} > 0$. The partitioning used for the derivation of (46) is: $\tilde{\xi}_1 = [\xi_1, \xi_2]$ and $\tilde{\xi}_2 = [\xi_3, \xi_4]$. The lower bound reads:

$$\begin{aligned} \mathbb{P}(T_i x \geq \xi_i, i = 1, \ldots, 4)\ &\geq\ \prod_{k=1}^{2} \mathbb{P}(\tilde{\xi}_k \in L_k) \\ &=\ \mathbb{P}(\xi_1 \leq T_1 x, \xi_2 \leq T_2 x)\ \mathbb{P}(\xi_3 \leq T_3 x, \xi_4 \leq T_4 x). \end{aligned}$$

**Corollary 3** *If the random variable* $\xi$ *has a lognormal density function, problem* **BS1**

$$\begin{aligned} \textbf{BS1}:\ &\min\ c^T x \\ &\textit{s. to}\ \ (1c)-(1d);\ \ (48) \end{aligned}$$

*is a convex approximation of* **SP1**. **BS1** is an inner approximation of **SP1** if (47) holds and is an outer approximation of **SP1** if (49) holds.

*Proof* The left side of (48) can be rewritten as a sum of logarithms of cumulative normal probability distributions as in (45). The convexity of the feasible area of (48) is then be justified as in Theorem 7. □

Observe that the partitioning of $\xi$ into subvectors satisfying the condition defined in Theorem 8 may not be unique. The identification of the possible partitions may not be obvious. Moreover, from a computational point of view, it is important to have subvectors of small size (i.e., $\leq$ 3) so that the corresponding multivariate probabilities appearing in the computation of the bounds (46) and (50) are easy to compute. In that perspective, we formulate below optimization problems that permits to form 1) a maximal cardinality partition of $\xi$ and 2) a partition with minimal size of the largest subvector.

Let $k \leq r$ be the index identifying subvector $\tilde{\xi}_k$. Let $\gamma_{ik}$ be a binary variable equal to 1 if $\xi_i$ is included into $\tilde{\xi}_k$ and equal to 0 otherwise, and $\beta_k, k = 1, \ldots, r$ be a variable taking value 1 if $\tilde{\xi}_k$ is not empty.

**Theorem 9** *An optimal solution* $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ *of the mixed-integer linear programming problem*

$$\max \sum_{k=1}^{r} \beta_k \tag{51a}$$

$$\text{s. to } \gamma_{ik} = \gamma_{jk} \quad i = 1, \ldots, r-1,\ j = i+1, \ldots, r : \rho_{ij} < 0,\ k = 1, \ldots, r \tag{51b}$$

$$\sum_{k=1}^{r} \gamma_{ik} = 1 \qquad i = 1, \ldots, r \tag{51c}$$

$$\beta_k \le \sum_{i=1}^{r} \gamma_{ik} \qquad k = 1, \ldots, r \tag{51d}$$

$$0 \le \beta_k \le 1 \qquad k = 1, \ldots, r \tag{51e}$$

$$\gamma_{ik} \in \{0, 1\} \qquad i = 1, \ldots, r, k = 1, \ldots, r, \tag{51f}$$

*defines a maximal cardinality partition of* $\xi$. *The number of subvectors is* $\sum_{k=1}^{r} \boldsymbol{\beta}_k$ *and the composition of the subvectors* $\tilde{\xi}_k$ *is defined by* $\boldsymbol{\gamma}_{ik}$.

*Proof* The constraints (51b) enforce the conditions (47) (see Theorem 8) enabling the derivation of the lower bound (46). The constraints (51c) guarantee the disjointedness of the subvectors by forcing each $\xi_i$ to be assigned to exactly one subvector $\tilde{\xi}_k$. The constraints (51d) and (51e) allow $\beta_k$ to take value at most 1 if subvector $\tilde{\xi}_k$ is not empty and (51a) makes sure to have $\beta_k = 1$ in the optimal solution when $\tilde{\xi}_k$ is not empty. Therefore, the objective function (51a) counts the number of subvectors and the optimal value gives the cardinality of the smallest partition. □

Note that each variable $\beta_k$ can be defined on [0, 1] (51e) and does not have to be defined as binary, due to (51a) and (51d).

Observe that there is a lot of symmetry (see Margot 2010) in the above problem, which complicates its solution. This tends to produce a number of equivalent, symmetric solutions by means of re-indexation. Take any arbitrary partitioning defined by the index sets $B_1 = \{1, 3\}$ and $B_2 = \{2, 4, 5\}$. It is equivalent to the one defined by $B_2 = \{1, 3\}$ and $B_1 = \{2, 4, 5\}$. We propose now a strengthened and equivalent version of the above mixed-integer linear problem which incorporates symmetry-breaking constraints (52a):

$$\max \sum_{k=1}^{r} \beta_k$$

$$\text{s. to } \gamma_{ik} = \gamma_{jk} \quad i = 1, \ldots, r-1,\ j = i+1, \ldots, r : \rho_{ij} < 0,\ k = 1, \ldots, i \tag{52a}$$

$$\sum_{k=1}^{i} \gamma_{ik} = 1 \qquad i = 1, \ldots, r \tag{52b}$$

$$\beta_k \le \sum_{i=k}^{r} \gamma_{ik} \qquad k = 1, \ldots, r \tag{52c}$$

$$\gamma_{ik} \in \{0, 1\} \qquad i = 1, \ldots, r, k = 1, \ldots, i, \quad (51e). \tag{52d}$$

Note how the symmetry-breaking constraints (52a) differ from (51b). They force, for example, $\xi_1$ to be assigned to subvector $\tilde{\xi}_1$, while $\xi_2$ can only be assigned to $\tilde{\xi}_1$ or $\tilde{\xi}_2$. The introduction of the symmetry-breaking constraints in turn allows for the reduction in the number of: (1) constraints (52a), (2) terms in the sum of (52b), (3) terms in the right-hand side of (52c), (4) binary variables, equal to $(r+1)r/2$ in (52d) instead of $r^2$ in (51f). We could also use the following precedence constraints:

$$\beta_k \geq \beta_{k+1}, \; k = 1, \ldots r-1.$$

Finally, we propose a last formulation in which the objective is to minimize the dimension of the largest subvector $\tilde{\xi}_k$. Let $r_k \in [0, r-k+1]$, $k = 1, \ldots, r$ be the dimension of $\tilde{\xi}_k$.

**Theorem 10** *An optimal solution* $(\mathbf{w}, \boldsymbol{\gamma})$ *of the mixed-integer linear programming problem:*

$$\min w \tag{53a}$$

$$s.\ to \;\; w \geq \sum_{i=k}^{r} \gamma_{ik} \qquad k = 1, \ldots, r \tag{53b}$$

$$(52a)-(52b);\ (52d).$$

*defines a partition in which the size* $\mathbf{w}$ *of the largest subvector is minimal. The composition of the subvectors* $\tilde{\xi}_k$ *is defined by* $\boldsymbol{\gamma}_{ik}$.

*Proof* Each $\gamma_{ik}$ defines if $\xi_i$ is part of $\tilde{\xi}_k$ due to (51b) and (51c). Therefore, $\sum_{i=k}^{r} \gamma_{ik}$ is the size of subvector $\tilde{\xi}_k$. It follows that minimizing $w$ coupled with the epigraph constraint (53b) provides a partition in which the largest subvector has minimal size. □

### 2.5 Bounding problems with sum of disjoint products

The approach called sum of disjoint products is based on the addition law of probabilities. It relies on the idea that if two or more events have nothing in common, the probability that at least one of them will occur is the sum of the probabilities of the separate events. The sum of disjoint products approach was first used to evaluate the reliability of a network (Abraham 1979) and the analysis of fault trees (Bennetts 1975). The functioning of a network typically requires all the components in a set to be operational jointly.

The sum of disjoint products approach computes the reliability of a network as the probability that all components in at least one of the minimal (component) sets functions properly (Heidtmann 2002). This involves the computation of the probability of a disjunctive normal form, since there exists several minimal sets and each is composed of a number of components (i.e., with Boolean-work or not-behavior) that must all work simultaneously.

Consider $r$ events $A_i, i = 1, \ldots, r$, and their complements $\overline{A}_i, i = 1, \ldots, r$. The union of the $r$ complemented events can be written as:

$$\bigcup_{i=1}^{r} \overline{A}_i = \underbrace{\overline{A}_1 A_2 \ldots A_r}_{V_1} \bigcup \underbrace{\overline{A}_2 A_3 \ldots A_r}_{V_2} \bigcup \ldots \bigcup \underbrace{\overline{A}_{r-1} A_r}_{V_{r-1}} \bigcup \underbrace{\overline{A}_r}_{V_r}. \tag{54}$$

All the sets $V_i, i = 1, \ldots, r$ in (54) are disjoint and constructed as the product of binary variables or events (*i.e.*, $A_i, \overline{A}_i$). Their disjointedness implies that the probability of their union is equal to the sum of their individual probabilities:

$$\mathbb{P}\left(\bigcup_{i=1}^{r} \overline{A}_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{r} V_i\right) = \sum_{i=1}^{r} \mathbb{P}(V_i). \tag{55}$$

This, combined with $\mathbb{P}\left(\bigcap_{i=1}^{r} A_i\right) = 1 - \mathbb{P}\left(\bigcup_{i=1}^{r} \overline{A}_i\right)$, implies that the probability of the intersection of $r$ events can therefore be computed as:

$$\mathbb{P}\left(\bigcap_{i=1}^{r} A_i\right) = 1 - \left(\mathbb{P}\left(\overline{A}_1, A_2, \ldots, A_r\right) + \mathbb{P}\left(\overline{A}_2, A_3, \ldots, A_r\right) + \mathbb{P}\left(\overline{A}_{r-1}, A_r\right) + \mathbb{P}\left(\overline{A}_r\right)\right). \tag{56}$$

We shall now use (56) to derive a second-order lower bound on the probability of the intersection of $r$ events. Observing that, for $\mathbb{P}(Tx \geq \xi) \geq p$ to hold, the probability of each uncomplemented event $A_i$ must be large (*i.e.*, $\geq p$) and the probability of each complemented event $\overline{A}_i$ is limited from above by $(1 - p)$, we conserve one uncomplemented event in each Boolean product $V_k, k = 1, \ldots, r - 1$ to obtain a lower bound on the probability of the intersection of the $r$ events:

$$\mathbb{P}\left(\bigcap_{i=1}^{r} A_i\right) \geq 1 - \left(\mathbb{P}\left(\overline{A}_1, A_2\right) + \mathbb{P}\left(\overline{A}_2, A_3\right) + \ldots + \mathbb{P}\left(\overline{A}_{r-1}, A_r\right) + \mathbb{P}\left(\overline{A}_r\right)\right). \tag{57}$$

Let $p_r$ and $q_i, i = 1, \ldots, r - 1$ be auxiliary decision variables that respectively represent the probabilities $\mathbb{P}\left(T_r x \geq \xi_r\right)$ and $\mathbb{P}\left(T_i x < \xi_i, T_{i+1} x \geq \xi_{i+1}\right), i = 1, \ldots, r - 1$. Theorem 11 follows.

**Theorem 11** *The nonlinear optimization problem*

$$\begin{aligned}
&\min\ c^T x \\
&s.\ to\ \ F_r(T_r x) \geq p_r && \text{(58a)}\\
&\qquad \mathbb{P}\left(T_i x < \xi_i, T_{i+1} x \geq \xi_{i+1}\right) \leq q_i \qquad i = 1, \ldots, r - 1 && \text{(58b)}\\
&\qquad p_r - \sum_{i=1}^{r-1} q_i \geq p && \text{(58c)}\\
&\qquad 0 \leq q_i \leq 1 - p \qquad i = 1, \ldots, r - 1 && \text{(58d)}\\
&\qquad p \leq p_r \leq 1 && \text{(58e)}\\
&\qquad (1c)\text{–}(1d) && \text{(58f)}
\end{aligned}$$

*is an inner approximation of* **SP1**. *Its optimal value is an upper bound on the optimal value of* **SP1**.

*Proof* Combined with (2),

$$\mathbb{P}(A_i, A_j) + \mathbb{P}(\overline{A}_i, A_j) + \mathbb{P}(A_i, \overline{A}_j) + \mathbb{P}(\overline{A}_i, \overline{A}_j) = 1, \tag{59}$$

requires $\mathbb{P}(\overline{A}_i, A_j) \leq 1 - p, i = 1, \ldots, r - 1, j = i + 1, \ldots, r$, which is enforced via (58b) and (58d). Constraint (2) also implies that the cumulative marginal probability of each uncomplemented event must be larger than or equal to $p$, which in ensured through (58a) and (58e). Finally, the lower bound on the probability of the intersection of $r$ events is always at most equal to the the left-hand side of (58c), which is at least equal to the right-hand side in (57), required to be $\geq p$. ∎

## 3 Numerical illustration

In this section, we illustrate the sharpness of the bounds presented in the previous sections. We use the STABIL problem (Prékopa et al. 1980), well-known in the stochastic programming literature. The problem involves the construction of a plan for the Hungarian electrical energy sector in the seventies. The model includes 71 decision variables, minimizes a linear cost objective function subject to deterministic constraints featuring requirements about manpower balance, investment features, foreign trade balance, balance of the state budget, finance, and electricity demand satisfaction as well as the following joint probabilistic constraint including four linear stochastic inequalities with random right-hand sides:

$$\mathbb{P}\left\{\begin{array}{c} -25x_{25} \geq \xi_1 \\ -16.67x_{26} \geq \xi_2 \\ 0.8696x_{24} + x_{40} \geq \xi_3 \\ 0.9(x_1 + x_2 + x_3 + x_4) - 0.115x_{24} \geq \xi_4 \end{array}\right\} \geq p \,. \tag{60}$$

We refer to Prékopa et al. (1980) for the detailed formulation of the linear inequalities. The objective function—to be minimized—is defined as the difference between the increase in the wage bill and the enterprise profit before taxation. The vector $\xi$ follows a multivariate normal distribution. The first two components of the four-dimensional random vector $\xi$ restrain the planned deficit of foreign trade (in \$US and roubles) to be below a certain level, while the last two components express the relationships between the electrical sector and the other sectors of the Hungarian economy. In Prékopa et al. (1980), $\xi$ has mean vector $\mu = [-48313,\ -426,\ 16000,\ 14950]$ and standard deviation vector $\sigma = [483,\ 4,\ 160, 190]$. The two correlation matrices $R_1$ and $R_2$ below were considered

$$R_1 = \begin{pmatrix} 1 & -0.8 & 0.4 & 0.4 \\ -0.8 & 1 & 0.1 & 0.1 \\ 0.4 & 0.1 & 1 & 0.9 \\ 0.4 & 0.1 & 0.9 & 1 \end{pmatrix}; \quad R_2 = \begin{pmatrix} 1 & -0.7 & 0.3 & 0.3 \\ -0.7 & 1 & 0.1 & 0.1 \\ 0.3 & 0.1 & 1 & 0.9 \\ 0.3 & 0.1 & 0.9 & 1 \end{pmatrix} \tag{61}$$

with probability levels $p = 0.9$ and 0.95.

We have formulated in AMPL and solved the convex formulations presented in the manuscript (true problem **SP1** and approximated problems **BO2**, **BO3**, **BH1**, **BI3**, and **BS1**) with the Ipopt solver complemented by a dynamic-link libray (DLL) computing the probability values and the gradient of the normal distributions. Table 1 presents the results.

Problem **BI3** is the only outer approximation. All approximated problem instances could be solved in less than 150 seconds. Note that (47) holds for **BS1**, which is thus an inner approximation problem whose optimal solution provides an upper bound. The partitioning of $\xi$ into subvectors has been implemented using the approach called maximal cardinality partition presented in Theorem 9: $\xi = (\tilde{\xi}_1, \tilde{\xi}_2, \tilde{\xi}_3)$ and $\tilde{\xi}_1 = (\xi_1, \xi_2)$, $\tilde{\xi}_2 = \xi_3$, $\tilde{\xi}_3 = \xi_4$.

While the differences in the optimal value of the true problem and these of the bounding problems are nor very marked, it is however noticeable that the LP approximated problem **BO3** is the most conservative and that the problem even becomes infeasible for one of the considered correlation matrices with $p$=0.95. The new proposed Hunter-type approximated problem **BH1** and the setwise bound approximation **BH1** provide tighter upper bounds than those provided by the Boole-Bonferroni first-order bounding problems **BO1–BO3**.

**Table 1** Comparison of solutions for true and approximated problems

| Formulations | $p$ | $R$ | Optimal value |
|---|---|---|---|
| SP1 | 0.9 | $R_1$ | −4370.31 |
| | 0.9 | $R_2$ | −4370.63 |
| | 0.95 | $R_1$ | −4369.42 |
| | 0.95 | $R_2$ | −4369.91 |
| BO1 & BO2 | 0.9 | Any $R$ | −4370.22 |
| | 0.95 | Any $R$ | −4369.11 |
| BO3 | 0.9 | Any $R$ | −4368.01 |
| | 0.95 | Any $R$ | Infeasible |
| BH1 | 0.9 | $R_1$ | −4370.25 |
| | 0.9 | $R_2$ | −4370.31 |
| | 0.95 | $R_1$ | −4369.16 |
| | 0.95 | $R_2$ | −4369.46 |
| BS1 | 0.9 | $R_1$ | −4370.22 |
| | 0.9 | $R_2$ | −4370.27 |
| | 0.95 | $R_1$ | −4369.19 |
| | 0.95 | $R_2$ | −4369.44 |
| BI3 | 0.9 | $R_1$ | −4370.44 |
| | 0.9 | $R_2$ | −4370.97 |
| | 0.95 | $R_1$ | −4369.65 |
| | 0.95 | $R_2$ | −4370.21 |

## 4 Conclusion

We consider probabilistically constrained stochastic programming problems with joint chance constraints and random right-hand sides. Such problems are particularly challenging for random variables with continuous distributions, because the calculation of the cumulative distribution function and its gradient values involves numerical integration and/or simulation in large dimensional spaces. We study various bounding schemes based on the Boole-Bonferroni inequalities, the binomial moment, product-type inequalities, and the setwise dependence concept that allow for the derivation of relaxations of probabilistic programming problems. We revisit known and provide new relaxations or bivariate cumulative distribution functions of smaller dimensional random vectors. We present the computational challenges posed by the relaxations, study their convexity, and identify conditions under which the relaxations are convex programming problems.

**Acknowledgements** M. Lejeune acknowledges the partial support of the Office of Naval Research [Grant N000141712420].

## References

Abraham, J. A. (1979). An improved algorithm for network reliability. *IEEE Transactions on Reliability*, *28*(1), 58–61.

Bagnoli, M., & Bergstrom, T. (2005). Log-concave probability and its applications. *Economic Theory*, *26*(2), 445–469.
Bennetts, R. G. (1975). On the analysis of fault trees. *IEEE Transactions on Reliability*, *24*(3), 175–185.
Charnes, A., Cooper, W. W., & Symonds, G. H. (1958). Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil. *Management Science*, *4*, 235–263.
Chen, W., Sim, M., Sun, J., & Teo, C.-P. (2010). From CVaR to uncertainty set: Implications in joint chance-constrained optimization. *Operations Research*, *58*(2), 470–485.
Chhetry, D., Kimeldorf, G., & Sampson, A. R. (1989). Concepts of setwise dependence. *Probability in the Engineering and Information Sciences*, *3*, 367–380.
Costigan, T. M. (1996). Combination Setwise–Bonferroni-type bounds. *Naval Research Logistics*, *43*, 59–77.
Deák, I. (1988). Multidimensional integration and stochastic programming. In Y. Ermoliev & R. J. B. Wets (Eds.), *Numerical techniques for stochastic optimization* (pp. 187–200). Berlin: Springer.
Deák, I. (2000). Subroutines for computing normal probabilities of sets—Computer experiences. *Annals of Operations Research*, *100*, 103–122.
Dentcheva, D., & Martinez, G. (2013). Regularization methods for optimization problems with probabilistic constraints. *Mathematical Programming*, *138*(1), 223–251.
Dentcheva, D., Prékopa, A., & Ruszczyński, A. (2000). Concavity and efficient points of discrete distributions in probabilistic programming. *Mathematical Programming*, *89*, 55–77.
Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, *1*, 141–149.
Genz, A., & Bretz, F. (2009). *Computation of multivariate normal and t probabilities. Lecture notes in statistics* (Vol. 195). Dordrecht: Springer.
Guigues, V., & Henrion, R. (2017). Joint dynamic probabilistic constraints with projected linear decision rules. *Optimization Methods & Software*, *32*(5), 1006–1032.
Heidtmann, K. D. (2002). Statistical comparison of two sum-of-disjoint-product algorithms for reliability and safety evaluations. In G. Goos, J. Hartmanis, & J. van Leeuwen (Eds.), *Computer safety, reliability and security—Lecture notes in computer science* (pp. 70–81). Berlin: Springer.
Henrion, R., & Möller, A. (2012). A gradient formula for linear chance constraints under Gaussian distribution. *Mathematics of Operations Research*, *37*, 475–488.
Hunter, D. (1976). Bounds for the probability of a union. *Journal of Applied Probability*, *13*, 597–603.
Kogan, A., & Lejeune, M. A. (2013). Threshold Boolean form for joint probabilistic constraints with random technology matrix. *Mathematical Programming*, *147*, 391–427.
Kwerel S. M. (1975). Bounds on the probability of the union and intersection of $m$ events. *Advances in Applied Probability*, 431–448.
Lejeune, M. A. (2012a). Pattern-based modeling and solution of probabilistically constrained optimization problems. *Operations Research*, *60*(6), 1356–1372.
Lejeune, M. A. (2012b). Pattern definition of the $p$-efficiency concept. *Annals of Operations Research*, *200*(1), 23–36.
Lejeune, M. A., & Margot, F. (2016). Solving chance-constrained optimization problems with stochastic quadratic inequalities. *Operations Research*, *64*(4), 939–957.
Lejeune, M. A., & Noyan, N. (2010). Mathematical programming generation of $p$-efficient points. *European Journal of Operational Research*, *207*(2), 590–600.
Lejeune, M. A., & Ruszczyński, A. (2007). An efficient trajectory method for probabilistic inventory-production-distribution problems. *Operations Research*, *55*(3), 1–17.
Margot, F. (2010). Symmetry in integer linear programming. In M. Junger, T. M. Liebling, D. Naddef, G. Nemhauser, W. R. Pulleyblank, G. Reinelt, G. Rinaldi, & L. A. Wolsey (Eds.), *50 years of integer programming* (pp. 647–686). Berlin: Springer.
Miller, B. L., & Wagner, H. M. (1965). Chance constrained programming with joint constraints. *Operations Research*, *13*, 930–945.
Mohtashami Borzadaran, G. R., & Mohtashami Borzadaran, H. A. (2011). Log-concavity property for some well-known distributions. *Surveys in Mathematics and its Applications*, *6*, 203–219.
Nemirovski, A., & Shapiro, A. (2006). Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, *17*(4), 969–996.
Prékopa A. (1970). On probabilistic constrained programming. In *Proceedings of the Princeton symposium on mathematical programming* (pp. 113–138). Princeton University Press.
Prékopa, A. (1973). Contributions to the theory of stochastic programming. *Mathematical Programming*, *4*, 202–221.
Prékopa, A. (1990). Dual method for a one-stage stochastic programming problem with random RHS obeying a discrete probability distribution. *Zeithchrift fur Operations Research*, *34*, 441–461.
Prékopa, A. (1995). *Stochastic programming*. Dordrecht-Boston: Kluwer.

Prékopa, A. (1999). The use of discrete moment bounds in probabilistic constrained stochastic programming models. *Annals of Operations Research*, *85*, 21–38.

Prékopa, A. (2001). On the concavity of multivariate probability distribution functions. *Operations Research Letters*, *29*, 1–4.

Prékopa, A. (2003). Probabilistic programming models. In A. Ruszczyński & A. Shapiro (Eds.), *Stochastic programming: Handbook in operations research and management science* (Vol. 10, pp. 267–351). Amsterdam: Elsevier Science.

Prékopa, A., Ganczer, S., Deák, I., & Patyi, K. (1980). The STABIL stochastic programming model and its experimental application to the electrical energy sector of the Hungarian economy. In M. A. Dempster (Ed.), *Stochastic programming* (pp. 369–385). London: Academic Press.

Ruszczynski, A. (2002). Probabilistic programming with discrete distribution and precedence constrained knapsack polyhedra. *Mathematical Programming*, *93*(2), 195–215.

Saxena, A., Goyal, V., & Lejeune, M. A. (2010). MIP reformulations of the probabilistic set covering problem. *Mathematical Programming*, *121*(1), 1–31.

Slepian, D. (1962). On the one-sided barrier problem for Gaussian noise. *Bell System Technical Journal*, *41*, 463–501.

Szántai, T. (1988). A computer code for solution of probabilistic-constrained stochastic programming problems. In Y. Ermoliev & R. J. B. Wets (Eds.), *Numerical techniques for stochastic optimization* (pp. 229–235). New York: Springer.

Wallace, S. W., & Ziemba, W. T. (2005). *Applications of stochastic programming*. Philadelphia, PA: MPS-SIAM Series on Optimization.

Worsley, K. J. (1982). An improved Bonferroni inequality and applications. *Biometrika*, *69*(2), 297–302.