Easy distributions for combinatorial optimization problems with probabilistic constraints

Bernard Fortz, Michael Poss

To cite this version:

HAL Id: hal-01255270
https://hal.science/hal-01255270
Submitted on 30 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Easy distributions for combinatorial optimization problems with probabilistic constraints

Bernard Fortz ∗ Michael Poss †

August 27, 2010

Abstract

We show how we can linearize individual probabilistic linear constraints with binary variables when all coefficients are independently distributed according to either $\mathcal{N}(\mu_i, \lambda \mu_i)$, for some $\lambda > 0$ and $\mu_i > 0$, or $\Gamma(k_i, \theta)$ for some $\theta > 0$ and $k_i > 0$. The constraint can also be linearized when the coefficients are independent and identically distributed and either positive or strictly stable random variables.

Keywords: Stochastic programming, combinatorial optimization, probabilistic constraint.

1 Introduction

Many combinatorial optimization models address problems with parameters which are impossible to predict exactly. Therefore, it is often more accurate to model these parameters with random variables. This modifies the structure of the optimization problems, depending on the times at which decisions are taken and parameters are revealed. In this note we study probabilistic constraints: all decisions must be taken here and now, such that the constraints of the model shall be satisfied with a certain probability. In other words, we aim at maximizing some objective for a given feasibility tolerance.

Stochastic programs with linear probabilistic constraints are in general non convex non linear optimization problems, see [18] among others. If furthermore some variables are integer, they become non convex Mixed Integer Non Linear Problems [15]. Although probabilistic constraints have been widely studied for many years, see [17] [31] [37] and the references therein, papers on problems with integer variables are not very numerous. Among them, problems featuring joint probabilistic constraints with random right hand side have been studied by [7] [8] [9] who propose exact and heuristic branch-and-bound algorithms, [13] who study formulations and bounding procedures, [23] who develop a column-generation based algorithm for a supply chain management problem, and [35] who introduce the concepts of p-inefficiency and provide extensive computational results for the probabilistic set-covering problem studied in [8]. All these works handle probabilistic constraints through the concept of p-efficient points introduced by [32], apart from [35] which uses p-inefficient points instead.

Herein, we consider problems where uncertainty affects both sides of the constraints. A branch-and-bound algorithm and heuristics for such problems have been proposed in [5] [6], and [21] studies valid inequalities for the problem with individual probabilistic constraints with uncertainty in both sides.

In what follows, we are particularly interested by the case of individual probabilistic constraints while the random variables follow particular continuous distributions, among which Gaussian distributions. Previous results in this direction assume that all random variables are normally distributed. In that case, the probabilistic constraints can be rewritten as quadratic constraints (see [19] [33] [39]), convex under some assumption on the confidence level [29]. If all variables are binary, the constraints can be further linearized using classical techniques [16]. Further work extends the classical gaussian framework to the more general class of radial distributions [12]. The authors show how a probabilistic constraint can be written as a second-order cone convex constraint. The latter constraint can be linearized as well when working with binary variables.

In this note, we always assume that coefficients are independent continuous random variables. We show that an individual linear probabilistic constraint with binary variables is equivalent to a linear constraint when all coefficients are distributed according to either $\mathcal{N}(\mu_i, \lambda \mu_i)$, for some $\lambda > 0$ and $\mu_i > 0$, or $\Gamma(k_i, \theta)$ for some $\theta > 0$ and $k_i > 0$. The constraint can also be linearized when the coefficients are independent and identically distributed, and either positive or strictly stable random variables.

The next section describes precisely the constraints studied herein. Then, in Section 3 we study the case of identically distributed random variables, while in Section 4 we study gaussian and gamma random variables. Finally, Section 5 illustrates our results on a multi-commodity flow problem arising in telecommunications networks, and we conclude in Section 6.
2 Studied constraints

In the following we study mainly the following type of probabilistic constraints,

\[ C_1(x) = P \left( \sum_{i=1}^{n} a_i x_i \leq b \right) \geq p, \]  

(1)

though our results extend easily to

\[ C_2(x) = P \left( \sum_{i=1}^{n} a_i x_i \leq c_1 y_1 + b \right) \geq p, \]  

(2)

and

\[ C_3(x) = P \left( \sum_{i=1}^{m} a_i x_i \leq \sum_{j=1}^{m} c_j y_j + b \right) \geq p \]  

(3)

where \( p \in (0, 1) \), \( a_i \) are independent random variables, \( c_i \)
and \( b \) are fixed coefficients. In addition, we always
consider that \( x_i, y_j \in \{0, 1\} \), for \( 1 \leq i \leq n \) and \( 1 \leq j \leq m \).
The first constraint \( [1] \) is the so-called knapsack
constraint, which plays an important role in capacitated
problems such as unsplittable multi-commodity flow and
weighted assignment problems. The second constraint
(2) appears when the choice of the capacitated facilities
to be built is part of the decision: \( b \) denotes the initial
capacity and \( c_1 \) the capacity provided by the facility.
Typical examples are network design and facility location
problems. Finally, in many technical problems we
must choose at most one out of a set of different facilities,
for instance, different capacities for a new link to install
in a telecommunication network. This is represented by
(3).

In what follows, we say that two constraints \( C_1(x) \geq 0 \)
and \( C_2(x) \geq 0 \) are equivalent, denoted by \( C_1(x) \geq 0 \iff \)
\( C_2(x) \geq 0 \), if the sets \( \{ x \in \{0, 1\}^n \text{ s.t. } C_1(x) \geq 0 \} \) and
\( \{ x \in \{0, 1\}^n \text{ s.t. } C_2(x) \geq 0 \} \) are equal.

3 Identically distributed variables

We first consider \( [1] \) for the simple example where \( a_i \) are
positive random variables identically distributed. Since
\( a_i \) are positive, we see that

\[ P \left( \sum_{i=1}^{m} a_i \leq b \right) \leq P \left( \sum_{i=1}^{m-1} a_i \leq b \right). \]  

(4)

Thus, the number of \( x_i \) that can be equal to 1 can cer-
tainly not exceed

\[ N(b) = \max_{1 \leq l \leq n} \left\{ l \text{ s.t. } P \left( \sum_{i=1}^{l} a_i \leq b \right) \geq p \right\}. \]  

(5)

Conversely, if some binary vector \( x \) satisfies \( \sum_{i=1}^{n} x_i \leq N(b) \), then certainly \( x \) satisfies \( [1] \) because \( a_i \) are identical-
distributed. Then, considering \( [2] \), the previous reasoning holds with \( N(b) \) for \( y_1 = 0 \), and with \( N(b+c_1) \) for \( y_1 = 1 \). Finally, this reasoning extends to the pair of constraints \( [3] \), since at most one of the \( y_j \) can be equal to 1. We just proved the following:

**Proposition 1.** Consider \( n \) independent identically dis-
tributed positive random variables \( a_i \), \( 1 \leq i \leq n \). Then,
for \( x_i, y_j \in \{0, 1\} \), \( 1 \leq i \leq n \) and \( 1 \leq j \leq m \), the
following constraints are equivalent:

1. \( C_1(x) \geq p \iff \sum_{i=1}^{n} x_i \leq N(b) \)
2. \( C_2(x) \geq p \iff \sum_{i=1}^{n} x_i \leq (N(b+c_1) - N(b))y_1 + N(b) \)
3. If furthermore, \( \sum_{j=1}^{m} y_j \leq 1 \), then \( C_3(x) \geq p \iff \sum_{i=1}^{n} x_i \leq \sum_{j=1}^{m} (N(b+c_j) - N(b))y_j + N(b) \)

with \( N(r) \) defined in \( [5] \) for any real \( r \).

In the following, we focus on results of type 1. since 2.
and 3. can be deduced from 1. by the above arguments.

Remark that computing the value of \( N(b) \) requires,
in general, the solution of a multivariate integral that
must be solved using efficient packages for numerical
integration, see [33]. For some distributions, this computa-
tional burden can be avoided. For instance, if all \( a_i \)
are uniformly distributed between 0 and 1, their sum is
distributed according to (see for instance [14])

\[ f(z) = \frac{1}{n!} \sum_{k=0}^{n+1} (-1)^k \binom{n+1}{k} [(z - k)_+]^n. \]

The uniform distributions are not the only distributions
which sum up nicely. Stable distributions satisfy interest-
ning summation properties too. Recall that if \( a_i \) are \( n \)
independent copies of a stable random variable \( a \), then
for any constants \( x_i \) the random variable \( \sum_{i=1}^{n} x_i a_i \) has
the same distribution as \( v_n a + w_n \) with some constants
\( v_n, w_n \) for some \( \alpha \in (0, 2] \), and \( w_n \). Moreover, \( a \)
is said strictly stable if \( w_n = 0 \) in the relation above.

For instance, the Levy distribution, with density function
equal to \( f(z; c) = \frac{1}{\sqrt{2\pi c z}} \exp\left(-\frac{z^2}{2c}\right) \) for \( z \geq 0 \), is positive
(satisfying the hypothesis of Proposition \( [1] \) and stable
so that sums of such distributions are easy to compute.
We refer to \( [27] \) for a good introduction to stable dis-
tributions.

In general, the support of stable distributions inter-
sects negative reals. For instance Gaussian and Cauchy
distributions always have negative tails. We show next
that property \( [4] \) still holds for strictly stable distribu-
tions. By definition

\[ \sum_{i=1}^{n} a_i \sim n^{1/\alpha} a_1 \quad \alpha \in (0, 2], \]
so that
\[ P \left( \sum_{i=1}^{n} a_i \leq b \right) = P(n^{1/\alpha} a_1 \leq b) = P(a_1 \leq bn^{-1/\alpha}). \]

If \( b \geq 0 \), the function \( n \mapsto bn^{-1/\alpha} \) is non increasing, implying (4). We obtain the following:

**Proposition 2.** Consider \( n \) independent identically distributed strictly stable random variables \( a_i \), \( 1 \leq i \leq n \), and \( b \geq 0 \). Then, if \( x_i \in \{0,1\} \) for each \( 1 \leq i \leq n \), the following constraints are equivalent:
\[ C_1(x) \geq p \iff \sum_{i=1}^{n} x_i \leq N(b), \]
with \( N(b) \) defined in (5).

An example of strictly stable distribution with \( \alpha = 1 \) is the Cauchy distribution, with density function
\[ f(z; x_0, \gamma) = \frac{1}{\pi} \left( \frac{\gamma}{(z-x_0)^2 + \gamma^2} \right) \]
for some location parameter \( x_0 \in \mathbb{R} \) and scale parameter \( \gamma > 0 \).

### 4 Non identically distributed variables

A well known stable distribution is the Gaussian distribution. In fact, for Gaussian and gamma random variables we are able to derive stronger results, allowing for the random variables to be distributed differently, as long as some regularity condition holds. Consider independent Gaussian random variables, \( a_i \sim N(\mu_i, \sigma_i^2) \), \( 1 \leq i \leq n \). Then, \( C_1(x) \geq p \) can be rewritten (see for instance [33])
\[ \sum_{i=1}^{n} \mu_i x_i + \Phi^{-1}(p) \sqrt{\sum_{i=1}^{n} \sigma_i^2 x_i^2} \leq b, \] (6)
where \( \Phi \) is the cumulative distribution of the standard Gaussian distribution \( N(0,1) \). When \( x \in \{0,1\}^n \), (6) can be linearized introducing additional continuous variables, see [10]. However, these linearizations contain significantly more variables than the direct linearization from Proposition 3 below.

**Proposition 3.** Consider \( n \) independent random variables \( a_i \sim N(\mu_i, \lambda_i) \), \( 1 \leq i \leq n \), for \( \lambda > 0 \) and \( \mu_i > 0 \). Then, if \( x_i \in \{0,1\} \) for each \( 1 \leq i \leq n \), the following constraints are equivalent:
\[ C_1(x) \geq p \iff \sum_{i=1}^{n} \mu_i x_i \leq \mu^*, \] (7)
where \( \mu^* \) is the unique root of the equation \( b - \mu = \Phi^{-1}(p) \lambda \mu \).

**Proof.** Gamma distributions satisfy also some kind of summation property, although weaker than the property satisfied by Gaussian distributions. Recall that if \( a_1, \ldots, a_n \) are independent Gamma with shape \( k_i \) and a common scale \( \theta \), then \( a := \sum_{i=1}^{n} a_i \sim \Gamma(k, \theta) \), with \( k = \sum_{i=1}^{n} k_i \). Thus, if \( x_i \) are binary numbers, we have also that \( a := \sum_{i=1}^{n} x_i a_i \sim \Gamma(k(x), \theta) \), with \( k(x) = \sum_{i=1}^{n} k_i x_i \). Thus, for binary input, \( C_1(x) \) is equivalent to \( P(\Gamma(k(x), \theta) \leq b) \) defined by
\[ P \left( \sum_{i=1}^{n} a_i \leq b \right) = \Phi \left( \frac{b - \mu(x)}{\sqrt{\lambda \mu(x)}} \right), \]
so that \( C_1(x) \geq p \) is equivalent to
\[ \frac{b - \mu(x)}{\sqrt{\lambda \mu(x)}} \geq \Phi^{-1}(p). \] (8)
The left hand side of (8) is decreasing in \( \mu(x) \), and thus \( C_1(x) \geq p \) is equivalent to \( \mu(x) \leq \mu^* \), where \( \mu^* \) is the unique root of the equation \( b - \mu = \Phi^{-1}(p) \lambda \mu \). \[ \square \]

We provide in Section 3 an application of Proposition 3 to a routing problem arising in telecommunications. Similar examples can be devised for the generalized assignment problem, see for instance the Proportional Mean-Variance Model from [38] which assumes that random variables are those from Proposition 3.

The next proposition considers the case of independent gamma random variables used, for instance, to model waiting and processing times in servers locations problems [10].

**Proposition 4.** Consider \( n \) independent random variables \( a_i \sim \Gamma(k_i, \theta) \), \( 1 \leq i \leq n \), for some \( \theta > 0 \) and \( k_i > 0 \), and assume that \( b > 0 \). Then, if \( x_i \in \{0,1\} \) for each \( 1 \leq i \leq n \), the following constraints are equivalent:
\[ C_1(x) \geq p \iff \sum_{i=1}^{n} k_i x_i \leq k^*, \]
where \( k^* \) is the unique solution of
\[ \int_{0}^{b} e^{z-k_1-x-k^2} \frac{dz}{\Gamma(k(x))} = p \]
and the gamma function is defined by
\[ \Gamma(k) = \frac{\int_{0}^{\infty} e^{z-k_1-x-k^2} \frac{dz}{\Gamma(k(x))}}{\Gamma(k(x))}. \]

**Proof.** Gamma distributions satisfy also some kind of summation property, although weaker than the property satisfied by Gaussian distributions. Recall that if \( a_1, \ldots, a_n \) are independent Gamma with shape \( k_i \) and a common scale \( \theta \), then \( a := \sum_{i=1}^{n} a_i \sim \Gamma(k, \theta) \), with \( k = \sum_{i=1}^{n} k_i \). Thus, if \( x_i \) are binary numbers, we have also that \( a := \sum_{i=1}^{n} x_i a_i \sim \Gamma(k(x), \theta) \), with \( k(x) = \sum_{i=1}^{n} k_i x_i \). Thus, for binary input, \( C_1(x) \) is equivalent to \( P(\Gamma(k(x), \theta) \leq b) \) defined by
\[ P \left( \sum_{i=1}^{n} a_i \leq b \right) = \Phi \left( \frac{b - \mu(x)}{\sqrt{\lambda \mu(x)}} \right), \]
so that \( C_1(x) \geq p \) is equivalent to
\[ \frac{b - \mu(x)}{\sqrt{\lambda \mu(x)}} \geq \Phi^{-1}(p). \] (8)
The left hand side of (8) is decreasing in \( \mu(x) \), and thus \( C_1(x) \geq p \) is equivalent to \( \mu(x) \leq \mu^* \), where \( \mu^* \) is the unique root of the equation \( b - \mu = \Phi^{-1}(p) \lambda \mu \). \[ \square \]
We are left to prove that $K(k)$ is a strictly decreasing function of $k > 0$:

$$\frac{dK}{dk}(k) = \frac{d}{dk} \int_0^b z^{k-1}e^{-z}dz = \frac{\theta}{\Gamma(k)} \int_0^b \ln(z)z^{k-1}e^{-z}dz \int_0^\infty v^{k-1}e^{-v}dv$$

$$- \int_0^b z^{k-1}e^{-z}dz \int_0^\infty \ln(v)v^{k-1}e^{-v}dv$$

$$= \frac{\theta}{\Gamma(k)} \int_0^b dz \int_0^\infty dv \left( z^{k-1}v^{k-1}e^{-z-v}\ln z \right),$$

which is strictly negative because $\ln \frac{z}{v} < 0$ for $(z,v) \in [0,b] \times (b,\infty)$.

When $b \leq 0$, $K(k(x)) = 0$ so that the probabilistic constraint is equivalent to $\sum_{i=1}^n x_i \leq 0$.

5 Application to the bandwidth packing problem

In what follows, we apply Proposition 3 to a multi-commodity flow problem occurring in telecommunications networks. We discuss different approaches to tackle the probabilistic constraints. Notice that our example is easily extended to the problem of designing a telecommunications network, replacing the fixed capacity $C$ by a set of facilities with capacity $C^j$, $j = 1, \ldots, m$ as in [3].

5.1 Problem description

Given a directed graph $G = (V,A)$ with a capacity vector $C$, and a set of commodities $K$ of size $d^k$ and revenue $c^k$ from $s(k)$ to $t(k)$ for each $k \in K$, the bandwidth packing problem (BWP) aims at routing commodities on the network in order to maximize the total revenue. For technical reasons based on routing protocols, each commodity must be sent along a unique path from $s(k)$ to $t(k)$, see [4, 20]. Introducing the binary variable $x^k_a$ stating whether commodity $k$ is routed through arc $a$, the problem can be formulated as

$$\max \sum_{k \in K} \sum_{a \in A^-}(s(k)) c^k_a x^k_a$$

s.t.

$$\sum_{a \in A^-(v)} x^k_a - \sum_{a \in A^+(v)} x^k_a = \begin{cases} 
1 & v = t(k) \\
-1 & v = s(k) \\
0 & \text{else}
\end{cases}$$

$$\sum_{k \in K} d^k_a x_a^k \leq C_a$$

$$x^k_a \in \{0,1\} \quad k \in K, a \in A,$$

where $A^+(v)$ and $A^-(v)$ denote the set of outgoing arcs and incoming arcs at node $v$, respectively. In practice, although the traffic size $d^k$ varies along time, it is not convenient to change the routing according the these variations: $x$ must be set once for a given time period. Different frameworks allow to model such uncertainties. Some works consider that $d$ belongs to a polyhedron $D$ and that $0$ must be feasible for any $d \in D$, see [22] and the closely related [2], among others. Others [31] model $d^k$, $k \in K$, by random variables and replace (9) by

$$P \left( \sum_{k \in K} d^k_a x^k_a \leq C_a \right) \geq p \quad a \in A. \quad (10)$$

In what follows, we assume that $d^k$, $k \in K$, are independent Gaussian distributed according to $N(\mu^k, \lambda^k \sigma^k)$. The Gaussian assumption has been studied in [11, 20] and used in [3], among others. Moreover, [36] (followed by [24] and [3]) assume that $d^k$ and $d^h$ are independently distributed for $k \neq h$. Finally, some authors [26] suggest that means and variances are linearly correlated as traffic size increases, that is, $\sigma = \lambda \mu$ for some $\lambda > 0$, so that we can apply Proposition 3 to (10).

5.2 Solution methods

We review different approaches to tackle the chance-constrained version of (BWP). Besides Proposition 3 there are two groups of methods to handle (10). Keeping the random vector continuous, we can tackle (10) by MINLP methods. Alternatively, we can sample the random variables to obtain a scenario set $S$ and solve the deterministic equivalent.

Direct linearization We apply Proposition 3 to (10), obtaining again problem (BWP) with $d^k$ and $C_a$ replaced by $\mu^k$ and the unique root $\mu^*_a$ of $C_a - \mu = \Phi^{-1}(p)\sqrt{\lambda^k}$, respectively. Computing $\mu^*_a$ is easy since function $C_a - \mu / \sqrt{\lambda^k}$ is convex and differentiable. Therefore, we can solve the problem with efficient algorithms used in the deterministic case, such as the branch-and-cut-and-price algorithm from [4].

MINLP methods When $p \geq 0.5$ and each $d^k$ is Gaussian, (10) is convex and thus, well suited for non linear algorithms [11]. However, it is clearly easier to use the direct linearization of (10) through Proposition 3 because non-linear constraints are harder to handle than linear ones and both formulations provide the same bound. For instance, outer approximation-based algorithms replace (10) by a set of tangent cutting planes. The latter contains more inequalities, with possibly highly fractional coefficients, than the unique inequality resulting from (10).

Alternatively, (10) with Gaussian random variables can be reformulated as (6). We can then rewrite (6)
as

\[
\sum_{k \in K} \mu^k x^k_a \leq C_a \quad (11)
\]

\[
\sum_{k \in K} \left[ (\Phi^{-1}(p)\sigma^k)^2 + \mu^k (2C_a - \mu^k) \right] x^k_a 
+ \sum_{k,h \in K : k \neq h} \left[ (\Phi^{-1}(p))^2 \sigma^k \sigma^h - 2\mu^k \mu^h \right] x^k_a x^h_a \leq C_a^2, \quad (12)
\]

for each \( a \in A \). When \( p > 0.5 \), which is the case in real situations, \( \mu^*_a < C_a \) and thus, (11) is less tight than (7). Hence, Proposition (10) allows to strengthen the above formulation by substituting (11) with (7). Then, (12) is not needed anymore to define a valid formulation. However, since it takes into account the binary restriction on \( x \) (by using \( (x^k_a)^2 = x^k_a \)), it may be used together with (7) to provide a stronger continuous relaxation. Note finally that linearizing (12) requires at least \( |K| \) additional variables and \( 2|K| \) additional constraints for each \( a \in A \) [16].

**Discretization and deterministic equivalent**

Sampling a scenario set \( S \) that approximates the continuous distribution \( d \) in an acceptable way, see [25, 28], among others, we can write a deterministic equivalent for (10):

\[
\sum_{k \in K} d^k x^k_a \leq C_a + M^a y_a, \quad a \in A, s \in S \quad (13)
\]

\[
\sum_{s \in S} p^s y^s_a \geq p, \quad a \in A, \quad (14)
\]

where components of vector \( M \) are numbers large enough. However, (13) and (14) yield a very difficult problem because (13) contains a large number of constraints and features “big-M” coefficients. Therefore, [5, 6] show how to replace (13) and (14) by a relevant set \( L \) of scenario sets through a branch-and-bound algorithm. Each \( l \in L \) yields a problem similar to (BWP), but with multiple capacity constraints [9] for each arc \( a \in A \) (one for each scenario in \( l \)). Then, using bounding mechanisms, they avoid solving all problems associated to elements of \( L \). Eventually, the exact approaches from [5, 6] will have solved several binary multi-commodity flow problems with multiple capacity constraints, each of them being more more complex than (BWP). Although applicable to a broader class of problems, this approach will in general be slower than the direct linearization from Proposition 3 that requires only to solve one problem similar to (BWP) plus the computation of the root vector \( \mu^* \).

**6 Conclusion**

In this note we show how to linearize individual probabilistic constraints with binary variables under very specific assumptions. Then, we propose a model for the bandwidth packing problem with probabilistic capacity constraints under demand uncertainty that satisfies our assumptions, and discuss different approaches to tackle the problem. We show that our linearization method simplifies considerably the stochastic problem because the latter becomes as easy to solve as its deterministic version. We can thus apply efficient algorithms taking the combinatorial structure of the problem into account.

In opposition, recent solution methods to integer programs featuring probabilistic constraints with uncertainty in both sides, such as [5, 6], are far more general and only assume that the uncertain parameters are described by a finite scenario set. Nevertheless, these approaches require to solve integer programs significantly harder than the deterministic versions.

**Acknowledgements**

This research is supported by an “Actions de Recherche Concertées” (ARC) projet of the “Communauté française de Belgique”. Michael Poss is a research fellow of the “Fonds pour la Formation à la Recherche dans l’Industrie et dans l’Agriculture” (FRIA). The authors would also like to thank a referee for constructive criticism.

**References**


