

# Optimization Models with Probabilistic Constraints

Darinka Dentcheva

Stevens Institute of Technology, Department of Mathematical Sciences  
Castle Point on Hudson, Hoboken, NJ 07030  
`ddentche@stevens-tech.edu`

**Summary.** This chapter presents an overview of the theory and numerical techniques for optimization models involving one or more constraints on probability functions. We focus on recent developments involving nonlinear probabilistic models. The theoretical fundament includes the theory and examples of generalized concavity for functions and measures, and some specific properties of probability distributions, including discrete distributions. We analyze the structure and properties of the constraining probabilistic functions and of the probabilistically constrained sets. An important part of the analysis is the development of algebraic constraints equivalent to the probabilistic ones. Optimality and duality theory for such models is presented.

In the overview of numerical methods for solving probabilistic optimization problems the emphasis is put on recent numerical methods for nonlinear probabilistically constrained problems based on the optimality and duality theory presented here. The methods provide optimal solutions for convex problems. Otherwise, they solve certain relaxations of the problem and result in suboptimal solutions and upper and lower bounds for the optimal value. Special attention is paid to probabilistic constraints with discrete distributions.

Some numerical approaches via statistical approximations are discussed as well. Numerical techniques of bounding probability in higher dimensional spaces with satisfactory precision are mentioned briefly in the context of discrete distributions. Application of combinatorial techniques in this context is sketched.

## 2.1 Introduction

Deterministic optimization models are usually formulated as problems of minimizing or maximizing a certain objective functional  $f(x)$  over  $x$  in a feasible set  $\mathcal{D}$  described by a finite system of inequalities

$$g_j(x) \leq 0, \quad j \in J,$$

with some functionals  $g_j$ ,  $j \in J$ .

When the objective functional or some of the constraint functionals depend not only on the decision vector  $x$ , but also on some random vector  $Z$ , the formulation of the optimization problem becomes unclear, and new precise definitions of the ‘objective’ and of the ‘feasible set’ are needed.

One way of dealing with that is to optimize the objective function and to require the satisfaction of the constraints *on average*. This leads to the following *stochastic optimization problem*:

$$\begin{aligned} & \min \mathbb{E}[f(x, Z)] \\ & \text{subject to } \mathbb{E}[g_j(x, Z)] \leq 0, \quad j \in J. \end{aligned}$$

We have assumed for this formulation that the expected value functions are well defined. More importantly, it assumes that the average performance is representative for our decision problem. When some of the quantities  $g_j(x, Z)$  have high variability a constraint on their expected value may not be satisfactory. When high uncertainty is involved another way to define the feasible set may be to impose constraints on probability functions, as in the following model:

$$\begin{aligned} & \min \mathbb{E}[f(x, Z)] \\ & \text{subject to } \mathbb{P}[g_j(x, Z) \leq 0, \quad j \in J] \geq p, \end{aligned} \tag{2.1}$$

where  $p \in (0, 1)$  is a modelling parameter expressing some fixed probability level. Constraints on probability are called *probabilistic* or *chance* constraints. The probability function can be formally understood as the expected value of the indicator function of the corresponding event. However, the discontinuity of the indicator function makes such problems qualitatively different than the expectation models.

In the following example probabilistic constraints arise in a natural way. Suppose we consider  $n$  investment opportunities, with random returns  $R_1, \dots, R_n$  in the next year. We have certain initial capital  $K$  and our aim is to invest some of it in such a way that the expected value of our investment after a year is maximized, under the condition that the chance of losing no more than a given fixed amount  $b > 0$  is at least  $p$ , where  $p \in (0, 1)$ . Such a requirement is called the *Value at Risk* (VaR) constraint.

Let  $x_1, \dots, x_n$  be the amounts invested in the  $n$  opportunities. Our investment changes in value after a year by  $g(x, R) = \sum_{i=1}^n R_i x_i$ . We can formulate the following stochastic optimization problem with probabilistic constraints:

$$\begin{aligned}
 & \max \sum_{i=1}^n \mathbb{E}[R_i]x_i \\
 & \text{s.t. } \mathbb{P} \left[ \sum_{i=1}^n R_i x_i \geq -b \right] \geq p \\
 & \sum_{i=1}^n x_i \leq K \\
 & x \geq 0.
 \end{aligned} \tag{2.2}$$

The constraint

$$\mathbb{P}[g_j(x, Z) \leq 0, j \in J] \geq p$$

is called *joint probabilistic constraint*, while the constraints

$$\mathbb{P}[g_j(x, Z) \leq 0] \geq p_j, j \in J, p_j \in [0, 1]$$

are called *individual probabilistic constraints*. Infinitely many individual probabilistic constraints appear naturally in the context of stochastic ordering constraints.

The notion of stochastic ordering or *stochastic dominance of first order* has been introduced in statistics in [203,212] and further applied and developed in economics [125,284]. It is defined as follows. For an integrable random variable  $X$  we consider its distribution function,  $F_X(\eta) = \mathbb{P}[X \leq \eta]$ ,  $\eta \in \mathbb{R}$ . We say that a random variable  $X$  *dominates in the first order a random variable*  $Y$  if

$$F_X(\eta) \leq F_Y(\eta) \quad \text{for all } \eta \in \mathbb{R}.$$

We denote this relation  $X \succeq_{(1)} Y$ . For two integrable random variables  $X$  and  $Y$ , we say that  $X$  *dominates  $Y$  in the second order* if

$$\int_{-\infty}^{\eta} F_X(\alpha) d\alpha \leq \int_{-\infty}^{\eta} F_Y(\alpha) d\alpha \quad \text{for all } \eta \in \mathbb{R}.$$

We denote this relation  $X \succeq_{(2)} Y$ . The second order dominance has been introduced in [150]. A modern perspective on stochastic ordering is presented in [232,237,352].

Returning to our example, we can require that the net profit on our investment dominates certain benchmark outcome  $Y$ , which may be the return of our current portfolio or some acceptable index. Then the VaR constraint has to be satisfied at a continuum of points. Setting  $\mathbb{P}[Y \leq \eta] = p_\eta$ , model (2.2) becomes:

$$\begin{aligned}
& \max \sum_{i=1}^n \mathbb{E}[R_i] x_i \\
& \text{s.t. } \mathbb{P} \left[ \sum_{i=1}^n R_i x_i \leq \eta \right] \leq p_\eta \text{ for all } \eta \in \mathbb{R} \\
& \sum_{i=1}^n x_i \leq K \\
& x \geq 0.
\end{aligned}$$

Using the stochastic dominance notation we can formulate the model as follows:

$$\begin{aligned}
& \max \sum_{i=1}^n \mathbb{E}[R_i] x_i \\
& \text{s.t. } \sum_{i=1}^n R_i x_i \succeq_{(1)} Y \\
& \sum_{i=1}^n x_i \leq K \\
& x \geq 0.
\end{aligned}$$

By changing the order of integration we can express the integrated distribution function as the expected shortfall: for each target value  $\eta$  we have

$$F_X^{(2)}(\eta) = \int_{-\infty}^{\eta} F_X(\alpha) d\alpha = \mathbb{E}[(\eta - X)_+],$$

where  $(\eta - X)_+ = \max(\eta - X, 0)$ . The integrated distribution function  $F_X^{(2)}(\cdot)$  is continuous, convex, non-negative and nondecreasing. It is well defined for all random variables  $X$  with finite expected value. A second order dominance constraint can be formulated as follows:

$$\sum_{i=1}^n R_i x_i \succeq_{(2)} Y \iff \mathbb{E}[(\eta - \sum_{i=1}^n R_i x_i)_+] \leq \mathbb{E}[(\eta - Y)_+] \text{ for all } \eta \in \mathbb{R}.$$

We can formulate the above model replacing the first order dominance constraint with the following constraints:

$$\mathbb{E}[(\eta - \sum_{i=1}^n R_i x_i)_+] \leq \mathbb{E}[(\eta - Y)_+] \text{ for all } \eta \in \mathbb{R}.$$

These second order dominance constraints can be viewed as a continuum of integrated chance constraints. In financial context it can be viewed as a continuum of *Conditional Value-at-Risk* (CVaR) constraints. For more information on this connection we refer to [110].

Models involving constraints on probability are introduced by Charnes *et al.* [81], Miller and Wagner [230], and Prékopa [275]. Problems with integrated chance constraints are considered in [152]. Models with stochastic dominance constraints are introduced and analyzed by Dentcheva and Ruszczyński in [107, 109, 111].

An essential contribution to the theory and solutions of problems with chance constraints was the theory of  $\alpha$ -concave measures and functions. In [276, 277] the concept of logarithmic concave measures is introduced and studied. This notion was generalized to  $\alpha$ -concave measures and functions in [51, 53, 63, 295], and further analyzed in [357], and [245]. Differentiability properties of probability functions are studied in [184, 185, 372, 373]. Statistical approximations of probabilistically constrained problems were analyzed by [178, 317]. For Monte Carlo approximations of chance constrained problems the reader is referred to [70, 71], see also Chapters 1 and 5 in this volume. Stability of models with probabilistic constraints is addressed in [103, 155–157, 303]. Nonlinear probabilistic problems are investigated in [104] where optimality conditions and duality results are established. Generalized concavity theory for discrete distributions and its consequences for probabilistic optimization is presented in [105, 106].

The formulation of the problem with probabilistic constraints is in harmony with the basic statistical principles used in testing statistical hypotheses and other statistical decisions. In engineering, reliability is frequently a central issue (*e.g.*, in telecommunication, transportation, hydrological network design and operation, engineering structure design, electronic manufacturing problems, *etc.*) and the problem with probabilistic constraints is very relevant. In finance, the concept of Value at Risk enjoys great popularity, [113, 300]. Integrated chance constraints represent a more general form of this concept. The concept of stochastic dominance is fundamental in economics and statistics (see [14, 107, 111, 132, 232]).

## 2.2 Structure and Properties of Probabilistically Constraint Sets

Fundamental questions to every optimization model concern the convexity of the feasible set, as well as continuity and differentiability of the constraint functions. The analysis of models with probability functions is based on specific properties of the underlying probability distributions. In particular, the *generalized concavity* theory plays a central role in probabilistic optimization. It facilitates the application of powerful tools of convex analysis.

### 2.2.1 Generalized Concavity of Functions and Measures

The generalized concavity discussed in this chapter is based on concavity of certain nonlinear transformation of the functions.

**Definition 1.** A non-negative function  $f(x)$  defined on a convex set  $D \subset \mathbb{R}^s$  is said to be  $\alpha$ -concave, where  $\alpha \in [-\infty, +\infty]$ , if for all  $x, y \in D$  and all  $\lambda \in [0, 1]$  the following holds:  
If  $\alpha = -\infty$  then

$$f(\lambda x + (1 - \lambda)y) \geq \min(f(x), f(y));$$

If  $\alpha = 0$  then

$$f(\lambda x + (1 - \lambda)y) \geq f^\lambda(x) f^{1-\lambda}(y);$$

If  $\alpha = \infty$  then

$$f(\lambda x + (1 - \lambda)y) \geq \max(f(x), f(y));$$

For any other value of  $\alpha$

$$f(\lambda x + (1 - \lambda)y) \geq [\lambda f^\alpha(x) + (1 - \lambda)f^\alpha(y)]^{1/\alpha}.$$

Here we take the following conventions:  $\ln 0 = -\infty$ ,  $0(+\infty) = 0$ ,  $0(-\infty) = 0$ ,  $0^0 = 1$ ,  $0^{-|\alpha|} = +\infty$ ,  $\infty^{-|\alpha|} = 0$ ,  $\infty^0 = 1$ .

In the case of  $\alpha = 0$  the function  $f$  is called *logarithmic concave*, and for  $\alpha = 1$  it is simply *concave*.

If  $f$  is  $\alpha$ -concave, then it is  $\beta$ -concave for all  $\beta \leq \alpha$ . Thus all  $\alpha$ -concave functions are  $(-\infty)$ -concave, that is, quasi-concave.

**Definition 2.** A probability measure  $\mathbb{P}$  defined on the Borel subsets of a convex set  $\Omega \subset \mathbb{R}^s$  is said to be  $\alpha$ -concave if for any Borel measurable subsets  $A$  and  $B$  of  $\Omega$  and for all  $\lambda$  we have the inequality

$$\mathbb{P}(\lambda A + (1 - \lambda)B) \geq \left( \lambda [\mathbb{P}(A)]^\alpha + (1 - \lambda) [\mathbb{P}(B)]^\alpha \right)^{1/\alpha},$$

where  $\lambda A + (1 - \lambda)B = \{\lambda x + (1 - \lambda)y : x \in A, y \in B\}$ . All special cases of  $\alpha$  and of one of the probabilities equal to 0 are treated as in Definition 1.

It is clear that if a random variable  $Z$  induces an  $\alpha$ -concave probability measure on  $\mathbb{R}^s$ , then its distribution function  $F_Z(x) = \mathbb{P}(Z \leq x)$  is an  $\alpha$ -concave function.

As usual, concavity properties imply certain continuity. The following theorem is due to Borell [53].

**Theorem 1.** If  $\mathbb{P}$  is a quasi-concave measure on  $\mathbb{R}^s$  and the dimension of its support is  $s$ , then  $\mathbb{P}$  has a density (with respect to the Lebesgue measure).

There is a relation between  $\alpha$ -concavity properties of measures and their densities (see [63, 280, 295] and references therein).

**Theorem 2.** Let  $\Omega$  be an open convex subset of  $\mathbb{R}^s$  and let  $m$  be the dimension of the smallest affine subspace  $L$  containing  $\Omega$ . The probability measure  $\mathbb{P}$  on  $\Omega$  is  $\gamma$ -concave with  $\gamma \in [-\infty, 1/m]$ , if and only if its probability density function with respect to the Lebesgue measure on  $L$  is  $\alpha$ -concave with

$$\alpha = \begin{cases} \gamma/(1 - m\gamma) & \text{if } \gamma < 1/m, \\ +\infty & \text{if } \gamma = 1/m. \end{cases}$$

**Corollary 1.** *Let an integrable non-negative function  $f(x)$  be defined on a non-degenerated convex set  $\Omega \subset \mathbb{R}^s$ . If  $f(x)$  is  $\alpha$ -concave with  $-1/s \leq \alpha \leq \infty$  and positive on the interior of  $\Omega$ , then the measure  $\mathbb{P}$  on  $\Omega$  defined as*

$$\mathbb{P}(A) = \int_A f(x) dx, \quad A \subset \Omega,$$

is  $\gamma$ -concave with

$$\gamma = \begin{cases} \alpha/(1 + s\alpha) & \text{if } \alpha \neq -1/s, \\ -\infty & \text{if } \alpha = -1/s. \end{cases}$$

The corollary states in particular that if a measure  $\mathbb{P}$  on  $\mathbb{R}^s$  has a density function  $f(x)$  such that  $f^{-1/s}$  is convex, then  $\mathbb{P}$  is quasi-concave.

For the following two results we refer the reader to [281].

**Theorem 3.** *If the  $s$ -dimensional random vector  $Z$  has a log-concave probability distribution and  $A$  is a constant  $m \times s$  matrix, then the  $m$ -dimensional random vector  $Y = AZ$  has a log-concave probability distribution.*

**Lemma 1.** *If  $\mathbb{P}$  is an  $\alpha$ -concave probability distribution and  $A \subset \mathbb{R}^s$  is a convex set, then the function  $f(x) = \mathbb{P}(A + x)$  is  $\alpha$ -concave.*

We extend the definition of generalized concavity to make it applicable to the case of discrete distributions. The first definition of discrete multivariate distributions is introduced in [29]. We adopt here the definition of [105] because it is more suitable to probabilistic optimization and it has essential consequences for optimality and duality theory of probabilistic optimization as it will become clear in Section 2.4.

**Definition 3.** *A distribution function  $F$  is called  $\alpha$ -concave on the set  $\mathcal{A} \subseteq \mathbb{R}^s$  with  $\alpha \in [-\infty, \infty]$ , if*

$$F(z) \geq (\lambda F(x)^\alpha + (1 - \lambda)F(y)^\alpha)^{1/\alpha}$$

for all  $z, x, y \in \mathcal{A}$  and  $\lambda \in (0, 1)$  such that  $z \geq \lambda x + (1 - \lambda)y$ . The special cases  $\alpha = 0$ , and  $\alpha = \pm\infty$  are treated the same way as in Definition 1.

Observe that if  $\mathcal{A} = \mathbb{R}^s$  this definition coincides with the usual definition of  $\alpha$ -concavity of a distribution function.

To illustrate the relation between Definition 1 and Definition 3 let us consider the case of integer random vectors which are roundups of continuously distributed random vectors. We denote the set of  $s$ -dimensional vectors with integer components by  $\mathbb{Z}^s$ .

*Remark 1.* If the distribution function of a random vector  $Z$  is  $\alpha$ -concave on  $\mathbb{R}^s$  then the distribution function of  $Y = \lceil Z \rceil$  is  $\alpha$ -concave on  $\mathbb{Z}^s$ .

The last property follows from the observation that at integer points both distribution functions coincide.

Furthermore for random vectors with independent components, we can relate the concavity of their marginal distributions to the concavity of the joint distribution.

**Theorem 4.** Assume that  $Z = (Z^1, \dots, Z^L)$ , where the  $s_l$ -dimensional sub-vectors  $Z_l$ ,  $l = 1, \dots, L$ , are independent ( $\sum_{l=1}^L s_l = s$ ). Furthermore, let the marginal distribution functions  $F_l : \mathbb{R}^{s_l} \rightarrow [0, 1]$  be  $\alpha_l$ -concave on sets  $\mathcal{A}_l \subset \mathbb{Z}^{s_l}$ .

1. If  $\alpha_l > 0$ ,  $l = 1, \dots, L$ , then  $F_Z$  is  $\alpha$ -concave on  $\mathcal{A} = \mathcal{A}_1 \times \dots \times \mathcal{A}_L$  with  $\alpha = (\sum_{l=1}^L \alpha_l^{-1})^{-1}$ ;
2. If  $\alpha_l = 0$ ,  $l = 1, \dots, L$ , then  $F_Z$  is log-concave on  $\mathcal{A} = \mathcal{A}_1 \times \dots \times \mathcal{A}_L$ .

For an integer random variable, our definition of  $\alpha$ -concavity is related to log-concavity of sequences.

**Definition 4.** A sequence  $p_k$ ,  $k \in \mathbb{Z}$ , is called log-concave, if

$$p_k^2 \geq p_{k-1}p_{k+1} \quad \text{for all } k \in \mathbb{Z}.$$

We have the following property (see [280, Theorem 4.7.2]):

**Theorem 5.** Suppose that for an integer random variable  $Y$  the probabilities  $p_k = \mathbb{P}\{Y = k\}$ ,  $k \in \mathbb{Z}$  form a log-concave sequence. Then the distribution function of  $Y$  is  $\alpha$ -concave on  $\mathbb{Z}$  for every  $\alpha \in [-\infty, 0]$ .

Another important property of  $\alpha$ -concave measures is the existence of a so-called floating body for all probability levels  $p \in [1/2, 1]$ . Let us recall that the support function of a convex set  $C \subset \mathbb{R}^s$  is defined as follows:

$$\sigma_C(h) = \sup\{\langle h, x \rangle : x \in C\}.$$

**Definition 5.** A measure  $\mathbb{P}$  on  $\mathbb{R}^s$  has a floating body for a level  $p > 0$  if there exists a convex set  $C_p \subset \mathbb{R}^s$  such that, for all vectors  $h \in \mathbb{R}^s$ ,

$$\mathbb{P}(\{x \in \mathbb{R}^s : \langle h, x \rangle \geq \sigma_{C_p}(h)\}) = p.$$

The set  $C_p$  is called the floating body of  $\mathbb{P}$  at level  $p$ .

All log-concave measures have a floating body, [225].

**Theorem 6.** Any log-concave probability measure has a floating body  $C_p$  for all levels  $p \in [1/2, 1]$ .



### 2.2.2 Examples of $\alpha$ -Concave Measures

1. The density of the *non-degenerate multivariate normal distribution* on  $\mathbb{R}^s$ :

$$f(x) = \frac{1}{\sqrt{(2\pi)^s \det \Sigma}} \exp \left( -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right),$$

where  $\Sigma$  is a positive definite matrix of dimension  $s \times s$  and  $\mu \in \mathbb{R}^s$ . Since the function  $\ln f(x)$  is concave (that is,  $f$  is 0-concave), the normal distribution is a log-concave measure.

2. The *uniform distribution on a convex set*  $D \subset \mathbb{R}^s$  with density

$$f(x) = \begin{cases} 1/V(D) & x \in D, \\ 0 & x \notin D, \end{cases}$$

where  $V(D)$  is the Lebesgue measure of  $D$ . The function  $f(x)$  is quasi-concave on  $D$ , hence it generates a  $1/s$ -concave measure on  $D$ .

3. The density function of the *multivariate Dirichlet's distribution* is defined as

$$f(x) = \begin{cases} \frac{\Gamma(\alpha_1 + \dots + \alpha_s)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_s)} x_1^{\alpha_1} x_2^{\alpha_2} \dots x_s^{\alpha_s} & \text{if } x_i \geq 0, \sum_i x_i = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Here  $\Gamma(\cdot)$  stands for the Gamma function. We define the open simplex

$$S = \left\{ x \in \mathbb{R}^s : \sum_{i=1}^s x_i = 1, x_i > 0, i = 1, \dots, s \right\}.$$

The function  $f(x)$  is  $(\alpha_1 + \dots + \alpha_s)^{-1}$ -concave on  $S$ , and therefore, the resulting measure is  $\beta$ -concave with  $\beta = (\alpha_1 + \dots + \alpha_s + s - 1)^{-1}$  on the closed simplex  $\bar{S}$ .

4. The density function of the  *$m$ -dimensional Student's distribution* with parameter  $n$

$$f(x) = \frac{\Gamma(\frac{m+n}{2}) \sqrt{\det A}}{\Gamma(\frac{n}{2}) \sqrt{(2\pi)^m}} \left( 1 + \frac{1}{n} (x - \mu)^T A (x - \mu) \right)^{-(m+n)/2},$$

where  $A$  is a positive definite matrix. Since  $f$  is  $(-\frac{2}{m+n})$ -concave, the corresponding measure is  $(-\frac{2}{n-m})$ -concave.

5. The density function of the  *$m$ -dimensional  $F$ -distribution* with parameters  $n_0, \dots, n_m$ , and  $n = \sum_{i=1}^m n_i$  is defined as follows:

$$f(x) = \text{const} \prod_{i=1}^m x_i^{n_i/2-1} \left( n_0 + \sum_{i=1}^m n_i x_i \right)^{-n/2}, \quad x_i \geq 0, \quad i = 1, \dots, m.$$

It is  $[-(n_0/2+m)^{-1}]$ -concave and the corresponding measure is  $(-\frac{2}{n})$ -concave.

6. The probability density function of the *Wishart distribution* is defined by

$$f(X) = \begin{cases} \frac{|X|^{\frac{N-q-2}{2}} e^{-\frac{1}{2} \text{tr } A^{-1} X}}{2^{\frac{N-1}{2}q} \pi^{\frac{q(q-1)}{4}} |A|^{\frac{N-1}{2}} \prod_{i=1}^q \Gamma\left(\frac{N-i}{2}\right)} & \text{for } X \succ 0 \\ 0 & \text{otherwise.} \end{cases}$$

Here  $X$  is assumed to be  $q \times q$  matrix containing the variables and  $A$  is fixed positive definite  $q \times q$  matrix. The symbol  $\succ$  denotes the partial order on the positive definite cone.

We assume that there are  $s = \frac{1}{2}q(q+1)$  independent variables and that  $N \geq q+2$ . The function  $f$  is log-concave.

7. The probability density function of the *beta distribution* is defined by

$$f(X) = \begin{cases} \frac{c(s_1, q)c(s_2, q)}{c(s_1 + s_2, q)} |X|^{\frac{1}{2}(s_1 - q - 1)} |I - X|^{\frac{1}{2}(s_2 - q - 1)} & \text{for } I \succ X \succ 0 \\ 0 & \text{otherwise.} \end{cases}$$

Here  $I$  stands for the identity matrix and the function  $c(\cdot, \cdot)$  is defined as follows:

$$\frac{1}{c(k, q)} = 2^{qk/2} \pi^{q(q-1)/2} \prod_{i=1}^q \Gamma\left(\frac{k-i+1}{2}\right).$$

We have assumed that  $s_1 \geq q+1$  and  $s_2 \geq q+1$ . The number of independent variables in  $X$  is  $s = \frac{1}{2}q(q+1)$ .

8. The *Cauchy distribution* regarded as a joint distribution of the random variables

$$Y_i = \frac{\sqrt{\nu} Z_i}{U} \quad i = 1, \dots, s,$$

where the random variables  $Z_1, \dots, Z_s$  have the standard normal distribution, each of them is independent of  $U$ , and  $U$  has the  $\chi$ -distribution with  $\nu$  degrees of freedom. The probability density function is

$$f(x) = \frac{\Gamma\left(\frac{1}{2}(\nu + s)\right)}{(\pi\nu)^{\frac{s}{2}} \Gamma\left(\frac{1}{2}\nu\right) |R|^{\frac{1}{2}}} \left(1 + \frac{1}{\nu} x^T R^{-1} x\right)^{-\frac{1}{2}(\nu + s)}$$

for  $x \in \mathbb{R}^s$ . If  $s = 1$  and  $\nu = 1$  this reduces to the well-known univariate Cauchy density

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad -\infty < x < \infty.$$

The  $s$ -variate Cauchy density has the property that  $f^{-\frac{1}{s}}$  is convex in  $\mathbb{R}^s$  and thus the distribution is quasi-concave by virtue of Corollary 1.

9. The probability density function of the *Pareto distribution* is

$$f(x) = a(a+1)\dots(a+s-1) \left( \prod_{j=1}^s \Theta_j \right)^{-1} \left( \sum_{j=1}^s \Theta_j^{-1} x_j - s + 1 \right)^{-(a+s)}$$

for  $x_i > \Theta_i$ ,  $i = 1, \dots, s$ , and  $f(x) = 0$  otherwise. Here  $\Theta_i$ ,  $i = 1, \dots, s$  are positive constants. Since  $f^{-\frac{1}{s}}$  is convex in  $\mathbb{R}^s$ , Corollary 1 implies that the Pareto distribution is quasi-concave.

10. A univariate *gamma distribution* is given by a probability density of the form

$$f(z) = \begin{cases} \frac{\lambda^\vartheta z^{\vartheta-1} e^{-\lambda z}}{\Gamma(\vartheta)} & \text{for } z > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Here  $\lambda > 0$  and  $\vartheta > 0$  are constants. For  $\lambda = 1$  the distribution is called the standard gamma distribution. If a random variable  $Y$  has the gamma distribution, then  $\lambda Y$  has the standard gamma distribution.

An  $s$ -variate gamma distribution can be defined by a certain linear transformation of  $s$  independent random variables  $Z_1, \dots, Z_s$  that have the standard gamma distribution. Given an  $s \times m$  matrix  $A$  ( $1 \leq m \leq 2^s - 1$ ) with 0-1 elements such that no column is the zero vector, setting  $Z = (Z_1, \dots, Z_s)$ , we define

$$Y = AZ.$$

The random vector  $Y$  has an  $s$ -variate standard gamma distribution. The univariate gamma density function is obviously log-concave. Thus, the  $s$ -variate standard gamma distribution is log-concave by virtue of Theorem 3.

11. Every distribution function of an  $s$ -dimensional *binary random vector* is  $\alpha$ -concave on  $\mathbb{Z}^s$  for all  $\alpha \in [-\infty, \infty]$ .

Indeed, let  $x, y$  be binary vectors,  $\lambda \in (0, 1)$  and let  $z \geq \lambda x + (1 - \lambda)y$ . Since  $z$  is integer and  $x$  and  $y$  binary, then  $z \geq x$  and  $z \geq y$ . Thus  $F(z) \geq \max(F(x), F(y)) = \max(F(x), F(y))$ . Consequently,  $F$  is  $\infty$ -concave.

12. The binomial, the Poisson, the geometric, and the hypergeometric one-dimensional probability distributions satisfy the conditions of Theorem 5 (see [280, p. 109]), and are, therefore, log-concave.

### 2.2.3 Convexity of Probabilistically Constrained Sets

Let us recall that a function  $g$  is called quasi-convex, if  $-g$  is quasi-concave in the sense of Definition 1. One of the most general results in the convexity theory of probabilistic optimization is the following:

**Theorem 7.** *Let  $g_j(\cdot, \cdot)$ ,  $j \in J$  be quasi-concave functions of the variables  $x \in \mathbb{R}^n$  and  $z \in \mathbb{R}^s$ . If  $Z \in \mathbb{R}^s$  is a random variable that has  $\alpha$ -concave probability distribution, then the function*

$$G(x) = \mathbb{P}[g_j(x, Z) \geq 0, j \in J]$$

*is  $\alpha$ -concave on the set*

$$D = \{x \in \mathbb{R}^n : \exists z \in \mathbb{R}^s \text{ such that } g_j(x, z) \geq 0, j \in J\}.$$

As a consequence, under the assumptions of Theorem 7, we obtain convexity statements for sets described by probabilistic constraints.

**Corollary 2.** *Assume that  $g_j(\cdot, \cdot)$ ,  $j \in J$  are quasi-concave functions jointly in both arguments, and that  $Z \in \mathbb{R}^s$  is a random variable that has an  $\alpha$ -concave probability distribution. Then the following set is convex and closed:*

$$\mathcal{X}_0 = \{x \in \mathbb{R}^n : \mathbb{P}[g_i(x, Z) \geq 0, i = 1, \dots, m] \geq p\}.$$

Observe that the closure of the set follows from the continuity of all  $\alpha$ -concave functions.

**Theorem 8.** *Given random variables  $Y_i \in \mathbb{R}^s$ , assume that  $g_j(\cdot, \cdot)$ ,  $j \in J$  are quasi-concave functions jointly in both arguments, and that  $Z_i$ ,  $i = 1, \dots, m$  have  $\alpha_i$ -concave distributions. Then the set with first order stochastic dominance constraint is convex and closed:*

$$\mathcal{X}_d = \{x \in \mathbb{R}^n : g_i(x, Z_i) \succeq_{(1)} Y_i, i = 1, \dots, m\}.$$

**Proof.** Let us fix  $i$  and  $\eta \in \mathbb{R}$  and consider the function

$$\mathbb{P}[g_i(x, Z_i) - \eta \leq 0] = 1 - \mathbb{P}[g_i(x, Z_i) - \eta > 0].$$

Constraint  $g_i(x, Z_i) \succeq_{(1)} Y_i$  can be formulated as follows:

$$\mathbb{P}[g_i(x, Z_i) - \eta > 0] \geq 1 - \mathbb{P}[Y_i \leq \eta] \quad \text{for all } \eta \in \mathbb{R}.$$

Denote the set of  $x$  satisfying this inequality by  $X_i(\eta)$ . By Theorem 7 the function at the left hand side of the last inequality is quasi-concave. Thus the set  $X_i(\eta)$  is convex and closed by Corollary 2. The set  $\mathcal{X}_d$  is the intersection of the sets  $X_i(\eta)$  for  $i = 1, \dots, m$  and all  $\eta \in \mathbb{R}$ , and, therefore, it is convex and closed.  $\square$

There is an intriguing relation between the sets constrained by first and second order dominance relation to a benchmark random variable (see [108]). We denote the space of integrable random variables by  $\mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$  and set

$$\begin{aligned} A_{(1)}(Y) &= \{X \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}) : X \succeq_{(1)} Y\}, \\ A_{(2)}(Y) &= \{X \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}) : X \succeq_{(2)} Y\}. \end{aligned}$$

It is proved in [107] that the set  $A_{(2)}(Y)$  is convex and closed in  $\mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ . The set  $A_{(1)}(Y)$  is closed, because convergence in  $\mathcal{L}_1$  implies convergence in probability. It is not convex in general.

**Theorem 9.** *Assume that  $Y$  has a continuous probability distribution function. Then*

$$A_{(2)}(Y) = \overline{\text{co}} A_{(1)}(Y),$$

where  $\overline{\text{co}} A_{(1)}(Y)$  stands for the closed convex hull of  $A_{(1)}(Y)$ .

If the underlying probability space is discrete and such that  $\Omega = \{1, \dots, N\}$ ,  $\mathcal{F}$  is the set of all subsets of  $\Omega$  and  $\mathbb{P}[k] = 1/N$ ,  $k = 1, \dots, N$ , we can remove the closure:

$$A_{(2)}(Y) = \text{co} A_{(1)}(Y),$$

Let us consider the special case

$$g_i(x, Z) := \langle a_i(Z), x \rangle + b_i(Z).$$

These functions are not necessarily quasi-concave in both arguments. If  $a_i(Z) = a_i$ ,  $i = 1, \dots, m$  we can apply Theorem 7 to conclude that the set  $\mathcal{X}_0$  is convex.

**Corollary 3.** *The following set is convex:*

$$\mathcal{X}_1 = \{x \in \mathbb{R}^n : \mathbb{P}[\langle a_i, x \rangle \leq b_i(Z), i = 1, \dots, m] \geq p\}$$

whenever  $b_i(\cdot)$  are quasi-concave functions and  $Z$  has a quasi-concave probability distribution.

If the functions  $g_i$  are not separable, we can invoke Theorem 6 (see also [202]).

**Corollary 4.** *The following set is convex:*

$$\mathcal{X}_1 = \{x \in \mathbb{R}^n : \mathbb{P}[\langle a_i(Z), x \rangle \leq b_i] \geq p_i, i = 1, \dots, m\} \quad (2.3)$$

whenever the vectors  $a_i(Z)$  have a log-concave probability distribution.

In particular, we obtain that the set  $\mathcal{X}_1$  is convex if  $a_i(Z)$  have one of the multivariate distributions from Section 2.2.2, e.g., the uniform, the normal, the Gamma distribution, etc.

### 2.2.4 Connectedness of Probabilistically Constrained Sets

It will be demonstrated later (Lemma 4) that the probabilistically constrained set  $\mathcal{X}$  is union of cones and, thus,  $\mathcal{X}$  could be disconnected. The following result provides a sufficient condition for  $\mathcal{X}$  to be topologically connected.

**Theorem 10.** *Assume that the functions  $g_i(\cdot, Z)$ ,  $i = 1, \dots, m$  are quasi-concave and that the following condition is satisfied: for all  $x^1, x^2 \in \mathbb{R}^n$  there exists a point  $x^* \in \mathbb{R}^n$  such that*

$$g_i(x^*, z) \geq \min\{g_i(x^1, z), g_i(x^2, z)\} \quad \text{for all } z \in \mathbb{R}^s, \text{ for all } i = 1, \dots, m.$$

*Then the set  $\mathcal{X}_0$  is connected.*

**Proof.** Let  $x^1, x^2 \in \mathcal{X}_0$  be arbitrary given points. We construct a path joining the two points, which is contained entirely in  $\mathcal{X}_0$ . Let  $x^*$  be the point that exists according to the assumption. We set

$$\pi(t) = \begin{cases} (1-2t)x^1 + 2tx^* & \text{for } 0 \leq t \leq 1/2 \\ 2(1-t)x^* + (2t-1)x^2 & \text{for } 1/2 < t \leq 1 \end{cases}$$

We observe that quasi-concavity of  $g_i$ ,  $i = 1, \dots, m$  and the assumptions of the theorem imply for  $0 \leq t \leq 1/2$  and for every  $i$  the following inequality:

$$g_i((1-2t)x^1 + 2tx^*, z) \geq \min\{g_i(x^1, z), g_i(x^*, z)\} = g_i(x^1, z).$$

Therefore,

$$\mathbb{P}[g(\pi(t)) \geq 0] \geq \mathbb{P}[g(x^1) \geq 0] \geq p \quad \text{for } 0 \leq t \leq 1/2.$$

Similar argument applies for  $1/2 < t \leq 1$ . Consequently,  $\pi(t) \in \mathcal{X}_0$ , and this proves the assertion.  $\square$

In [155] a slightly more general version of this result is proved in order to deal with probabilistic constraints involving stochastic processes.

## 2.3 Random Right Hand Side

We pay special attention to problems with separable constraint functions. Consider the following problem:

$$\begin{aligned} & \max f(x) \\ & \text{subject to } \mathbb{P}[g(x) \geq Z] \geq p, \\ & \quad x \in \mathcal{D}. \end{aligned} \tag{2.4}$$

Assume that  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $i = 1, \dots, m$ , are concave functions. Let  $\mathcal{D} \subseteq \mathbb{R}^n$  be a closed convex set, and  $Z$  be an  $m$ -dimensional random vector. We denote  $g = (g_1, \dots, g_m)$ . For two vectors  $a$  and  $b$  the inequality  $a \leq b$  is understood componentwise.

### 2.3.1 Continuity and Differentiability Properties of Distribution Functions

When the probabilistic constraint involves inequalities with random variables on the right hand side only, we can express it as a constraint on a distribution function:

$$\mathbb{P}[g(x) \geq Z] \geq p \iff F_Z(g(x)) \geq p$$

Thus, continuity and differentiability properties of distribution functions are relevant to the numerical solution of probabilistic optimization problems.

**Theorem 11.** *Suppose that  $Z$  has an  $\alpha$ -concave distribution with  $\alpha \in (-\infty, 0)$  and that the support of it  $\text{supp}P_Z$  has non-empty interior in  $\mathbb{R}^s$ . Then  $F_Z$  is locally Lipschitz-continuous on  $\text{int supp}P_Z$ .*

**Proof.** From the assumption that  $P_Z$  is  $\alpha$ -concave for  $\alpha < 0$ , we infer that the function  $F_Z^\alpha(\cdot)$  is convex. The assertion follows from the fact that convex functions are locally Lipschitz on the interior of their domain and from the Lipschitz continuity of the mapping  $t \mapsto t^{1/\alpha}$  away from 0. Fixing a point  $z \in \text{int supp}P_Z$ , there is a neighbourhood  $U$  of  $z$  contained in the interior of the support and such that  $F_Z^\alpha$  is locally Lipschitz with Lipschitz constant  $L_1$ . Decreasing  $U$  if necessary, we can find a compact set  $K$  such that  $U \subset K \subset \text{int supp}P_Z$ . Thus,  $\min_{z \in K} F_Z(z) = r > 0$ . Let  $L_2$  be the Lipschitz constant of  $t \mapsto t^{1/\alpha}$  on the interval  $[r, 1]$ . We obtain

$$|F_Z(z_1) - F_Z(z_2)| \leq L_2 |F_Z^\alpha(z_1) - F_Z^\alpha(z_2)| \leq L_2 L_1 \|z_1 - z_2\|.$$

□

**Theorem 12.** *Suppose that all one-dimensional marginal distribution functions of an  $s$ -dimensional random vector  $Z$  are locally Lipschitz continuous. Then  $F_Z$  is locally Lipschitz-continuous as well.*

**Proof.** The statement can be proved by straightforward evaluation of the distribution function by marginals for  $s = 2$  and mathematical induction on the dimension of the space. □

Assume that the measure  $P_Z$  has a density. It should be emphasized that the continuity and the essential boundedness of the density do not imply the Lipschitz continuity of the distribution function  $F_Z$ .

**Theorem 13.** *Assume that  $P_Z$  has a continuous density  $\theta(\cdot)$  and that all one-dimensional marginal densities are continuous as well. Then  $F_Z$  is continuously differentiable.*

**Proof.** We demonstrate the statement for  $s = 2$ . The assertions then follow by induction. The existence of the partial derivatives follows from the continuity of the density  $\theta$  by virtue of the theorem of Fubini:

$$\frac{\partial F_x}{\partial z_1}(z_1, z_2) = \int_{-\infty}^{z_2} \theta(z_1, t) dt \quad \text{and} \quad \frac{\partial F_x}{\partial z_2}(z_1, z_2) = \int_{-\infty}^{z_1} \theta(t, z_2) dt.$$

First, we observe that the mapping  $(x_1, x_2) \mapsto \int_a^{x_2} \theta(x_1, t) dt$  is continuous for every  $a \in \mathbb{R}$  by the uniform continuity of  $\theta(\cdot)$  on compact sets in  $\mathbb{R}^2$ . Given the points  $(x_1, x_2)$  and  $(y_1, y_2)$ , we have:

$$\begin{aligned} \left| \frac{\partial F}{\partial x_1}(x_1, x_2) - \frac{\partial F}{\partial y_1}(y_1, y_2) \right| &= \left| \int_{-\infty}^{x_2} \theta(x_1, t) dt - \int_{-\infty}^{y_2} \theta(y_1, t) dt \right| \\ &\leq \left| \int_{x_2}^{y_2} \theta(y_1, t) dt \right| + \left| \int_{-\infty}^{x_2} [\theta(x_1, t) - \theta(y_1, t)] dt \right| \leq \varepsilon. \end{aligned}$$

The last inequality is satisfied if the points  $(x_1, x_2)$  and  $(y_1, y_2)$  are sufficiently close by the continuity of  $\theta(\cdot)$  and the uniform continuity of the function  $(x_1, x_2) \mapsto \int_a^{x_2} \theta(x_1, t) dt$ .

The limit exists uniformly around  $x_1$  because of the continuity of the one-dimensional marginal densities.  $\square$

### 2.3.2 $p$ -Efficient Points

We concentrate on deriving an equivalent algebraic description of the feasible set. The level set of the distribution function of  $Z$  can be described as follows:

$$\mathcal{Z} = \{z \in \mathbb{R}^m : \mathbb{P}[Z \leq z] \geq p\}. \quad (2.5)$$

Clearly, problem (2.4) can be compactly rewritten as

$$\begin{aligned} &\max f(x) \\ &\text{subject to } g(x) \in \mathcal{Z}, \\ &x \in \mathcal{D}. \end{aligned} \quad (2.6)$$

**Lemma 2.** *For every  $p \in (0, 1)$  the level set  $\mathcal{Z}$  is non-empty and closed.*

**Proof.** The assertion follows from the monotonicity and the right continuity of the distribution function.  $\square$

Till the end of this section we denote the probability distribution function of  $Z$  by  $F$  omitting the subscript. The marginal probability distribution function of the  $i$ th component  $Z_i$  will be denoted by  $F_i$ .

We recall the concept of a  $p$ -efficient point.

**Definition 6.** *Let  $p \in (0, 1]$ . A point  $v \in \mathbb{R}^m$  is called a  $p$ -efficient point of the probability distribution function  $F$ , if  $F(v) \geq p$  and there is no  $z \leq v$ ,  $z \neq v$  such that  $F(z) \geq p$ .*



The  $p$ -efficient points are minimal points of the level set  $\mathcal{Z}$  with respect to the partial order in  $\mathbb{R}^m$  generated by the non-negative cone. This notion was first introduced in [278]. Similar concept is used in [326]. The concept was studied and applied in the context of discrete distributions and linear problems in the papers [105, 106, 282] and in the context of general distributions in [104].

Obviously, for a scalar random variable  $Z$  and for every  $p \in (0, 1]$  there is exactly one  $p$ -efficient point: the smallest  $v$  such that  $F(v) \geq p$ . Since  $F(v) \leq F_i(v_i)$  for every  $v \in \mathbb{R}^m$  and  $i = 1, \dots, m$ , we obtain that the set of  $p$ -efficient points is bounded from below.

**Lemma 3.** *Let  $p \in (0, 1]$  and let  $l_i$  be the  $p$ -efficient point of the one-dimensional marginal distribution  $F_i$ ,  $i = 1, \dots, m$ . Then every  $v \in \mathbb{R}^m$  such that  $F(v) \geq p$  must satisfy the inequality  $v \geq l = (l_1, \dots, l_m)$ .*

Let  $p \in (0, 1)$  and let  $v^j$ ,  $j \in J$ , be all  $p$ -efficient points of  $Z$ , where  $J$  is an arbitrary set. Denoting the positive orthant in  $\mathbb{R}^m$  by  $\mathbb{R}_+^m$ , we define the cones

$$K_j = v^j + \mathbb{R}_+^m, \quad j \in J.$$

The following result can be derived from Phelps theorem [267, Lemma 3.12] about the existence of conical support points, but an easy direct proof is provided.

**Theorem 14.**  $\mathcal{Z} = \bigcup_{j \in J} K_j$ .

**Proof.** If  $y \in \mathcal{Z}$  then either  $y$  is  $p$ -efficient or there exists a vector  $w$  such that  $w \leq y$ ,  $w \neq y$ ,  $w \in \mathcal{Z}$ . By Lemma 3, one must have  $l \leq w \leq y$ . The set  $Z_1 := \{z \in \mathcal{Z} : l \leq z \leq y\}$  is compact because the set  $\mathcal{Z}$  is closed. Thus, there exists  $w^1 \in Z_1$  with the minimal first coordinate. If  $w^1$  is a  $p$ -efficient point, then  $y \in w^1 + \mathbb{R}_+^m$ , what had to be shown. Otherwise, we define  $Z_2 := \{z \in \mathcal{Z} : l \leq z \leq w^1\}$ , and choose a point  $w^2 \in Z_2$  with the minimal second coordinate. Proceeding in the same way, we shall find the minimal element  $w^m$  in the set  $\mathcal{Z}$  with  $w^m \leq w^{m-1} \leq \dots \leq y$ . Therefore,  $y \in w^m + \mathbb{R}_+^m$ , and this completes the proof.  $\square$

By virtue of Theorem 14 we obtain (for  $0 < p < 1$ ) the following *disjunctive semi-infinite* formulation of problem (2.6):

$$\begin{aligned} & \max f(x) \\ & \text{subject to } g(x) \in \bigcup_{j \in J} K_j, \\ & x \in \mathcal{D}. \end{aligned} \tag{2.7}$$

Its main advantage is an insight into the nature of the non-convexity of the feasible set. The main difficulty is the implicit character of the disjunctive constraint.

Let  $S$  stand for the simplex in  $\mathbb{R}^{m+1}$ ,  $S = \{\alpha \in \mathbb{R}^{m+1} : \sum_{i=1}^{m+1} \alpha_i = 1, \alpha_i \geq 0\}$ . We define the convex hull of the  $p$ -efficient points:

$$E = \left\{ \sum_{i=1}^{m+1} \alpha_i v^{j_i} : \alpha \in S, j_i \in J \right\}.$$

The convex hull of  $\mathcal{Z}$  has a semi-infinite disjunctive representation as well.

**Lemma 4.**  $\text{co } \mathcal{Z} = E + \mathbb{R}_+^m$ .

**Proof.** By Theorem 14 every point  $y \in \text{co } \mathcal{Z}$  can be represented as a convex combination of points in the cones  $K_j$ . By the theorem of Caratheodory we can write  $y = \sum_{i=1}^{m+1} \alpha_i (v^{j_i} + w^i)$ , where  $w^i \in \mathbb{R}_+^m$ ,  $\alpha \in S$  and  $j_i \in J$ . The vector  $w = \sum_{i=1}^{m+1} \alpha_i w^i$  belongs to  $\mathbb{R}_+^m$ . Therefore,  $y \in \sum_{i=1}^{m+1} \alpha_i v^{j_i} + \mathbb{R}_+^m$ .  $\square$

**Theorem 15.** For every  $p \in (0, 1)$  the set  $\text{co } \mathcal{Z}$  is closed.

**Proof.** Consider a sequence  $\{z^k\}$  of points of  $\text{co } \mathcal{Z}$  which is convergent to a point  $\bar{z}$ . We have

$$z^k = \sum_{i=1}^{m+1} \alpha_i^k y_i^k,$$

with  $y_i^k \in \mathcal{Z}$ ,  $\alpha_i^k \geq 0$ , and  $\sum_{i=1}^{m+1} \alpha_i^k = 1$ . By passing to a subsequence, if necessary, we can assume that the limits

$$\bar{\alpha}_i = \lim_{k \rightarrow \infty} \alpha_i^k$$

exist for all  $i = 1, \dots, m+1$ . By Lemma 3 all points  $y_i^k$  are bounded below by some vector  $l$ . For simplicity of notation we may assume that  $l = 0$ .

Let  $I = \{i : \bar{\alpha}_i > 0\}$ . Clearly,  $\sum_{i \in I} \bar{\alpha}_i = 1$ . We obtain

$$z^k \geq \sum_{i \in I} \alpha_i^k y_i^k.$$

We observe that  $0 \leq \alpha_i^k y_i^k \leq z^k$  for all  $i \in I$  and all  $k$ . Since  $\{z^k\}$  is convergent and  $\alpha_i^k \rightarrow \bar{\alpha}_i > 0$ , each sequence  $\{y_i^k\}$ ,  $i \in I$ , is bounded. Therefore we can assume that each of them is convergent to some limit  $\bar{y}_i$ ,  $i \in I$ . By virtue of Lemma 2  $\bar{y}_i \in \mathcal{Z}$ . Passing to the limit in the last displayed inequality we obtain

$$\bar{z} \geq \sum_{i \in I} \bar{\alpha}_i \bar{y}_i \in \text{co } \mathcal{Z}.$$

Due to Lemma 4,  $\bar{z} \in \text{co } \mathcal{Z}$ .  $\square$

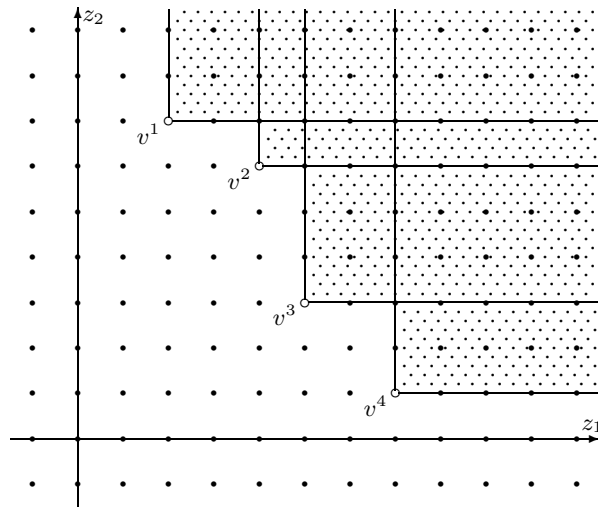
**Theorem 16.** For every  $p \in (0, 1)$  the set of extreme points of  $\text{co } \mathcal{Z}$  is non-empty and it is contained in the set of  $p$ -efficient points.

**Proof.** The set  $\text{co } \mathcal{Z}$  is included in  $l + \mathbb{R}_+^m$ , by virtue of Lemma 3 and Lemma 4. Therefore, it does not contain any line. Since it is closed by Theorem 15, it has at least one extreme point.

Let  $w$  be an extreme point of  $\text{co } \mathcal{Z}$ . Suppose that  $w$  is not a  $p$ -efficient point. Then Theorem 14 implies that there exists a  $p$ -efficient point  $v \leq w$ ,  $v \neq w$ . Since  $w + \mathbb{R}_+^m \subset \text{co } \mathcal{Z}$ , the point  $w$  is a convex combination of  $v$  and  $w + (w - v)$ . Consequently,  $w$  cannot be extreme.  $\square$

For a general random vector the set of  $p$ -efficient points may be unbounded and not closed.

The representation becomes very handy for problem (2.26) when the vector  $Z$  has a discrete distribution on  $\mathbb{Z}^s$ . In [105] discrete distributions are investigated, where the random vector  $Z$  takes values on a grid. Without loss of generality we can assume that  $Z \in \mathbb{Z}^s$ . Figure 2.1 illustrates the structure of the probabilistically constrained set.



**Figure 2.1.** Example of the set  $\mathcal{Z}$  with  $p$ -efficient points  $v^1, \dots, v^4$

**Theorem 17.** *For each  $p \in (0, 1)$  the set of  $p$ -efficient points of an integer random vector is non-empty and finite.*

**Proof.** The result follows from Dickson's Lemma [32, Corollary 4.48] and Lemma 3. For convenience we provide a short proof here. We shall at first show that at least one  $p$ -efficient point exists. Since  $p < 1$ , there must exist  $y$  such that  $F(y) \geq p$ . By Lemma 3, all  $v$  such that  $F(v) \geq p$  are bounded below by the vector  $l$  of  $p$ -efficient points of one-dimensional marginals. Therefore, if  $y$  is not  $p$ -efficient, one of finitely many integer points  $v$  such that  $l \leq v \leq y$  must be  $p$ -efficient.

Now we prove the finiteness of the set of  $p$ -efficient points. Suppose that there exists an infinite sequence of different  $p$ -efficient points  $v^j$ ,  $j = 1, 2, \dots$

Since they are integer, and the first coordinate  $v_1^j$  is bounded from below by  $l_1$ , with no loss of generality we may select a subsequence which is non-decreasing in the first coordinate. By a similar token, we can select further subsequences which are non-decreasing in the first  $k$  coordinates ( $k = 1, \dots, s$ ). Since the dimension  $s$  is finite, we obtain a subsequence of different  $p$ -efficient points which is non-decreasing in all coordinates. This contradicts the definition of a  $p$ -efficient point.  $\square$

Our proof can be easily adapted to the case of non-uniform grids for which a uniform lower bound on the distance of grid points in each coordinate exists. In this way we obtain the following *disjunctive* formulation with a finite index set  $J$  for problem (2.6):

$$\begin{aligned} & \min f(x) \\ & \text{subject to } g(x) \in \bigcup_{j \in J} K_j, \\ & \quad x \in \mathcal{D}. \end{aligned} \tag{2.8}$$

The concept of  $\alpha$ -concavity on a set can be used at this moment to find an equivalent representation of the set  $\mathcal{Z}$  for the discrete distributions.

**Theorem 18.** *Let  $\mathcal{Z}$  be the set of all possible values of an integer random vector  $Z$ . If the distribution function  $F$  of  $Z$  is  $\alpha$ -concave on  $\mathcal{Z} + \mathbb{Z}_+^s$ , for some  $\alpha \in [-\infty, \infty]$ , then for every  $p \in (0, 1)$  one has*

$$\mathcal{Z} = \{y \in \mathbb{R}^s : y \geq z \geq \sum_{j \in J} \lambda_j v^j, \sum_{j \in J} \lambda_j = 1, \lambda_j \geq 0, z \in \mathbb{Z}^s\},$$

where  $v^j$ ,  $j \in J$ , are the  $p$ -efficient points of  $F$ .

**Proof.** By the monotonicity of  $F$  we have  $F(y) \geq F(z)$  if  $y \geq z$ . It is, therefore, sufficient to show that  $\mathbb{P}(Z \leq z) \geq p$  for all  $z \in \mathbb{Z}^s$  such that  $z \geq \sum_{j \in J} \lambda_j v^j$  with  $\lambda_j \geq 0$ ,  $\sum_{j \in J} \lambda_j = 1$ . We consider five cases with respect to  $\alpha$ .

*Case 1:*  $\alpha = \infty$ . It follows from the definition of  $\alpha$ -concavity that

$$F(z) \geq \max\{F(v^j), j \in J : \lambda_j \neq 0\} \geq p.$$

*Case 2:*  $\alpha = -\infty$ . Since  $F(v^j) \geq p$  for each index  $j \in J$  such that  $\lambda_j \neq 0$ , the assertion follows as in Case 1.

*Case 3:*  $\alpha = 0$ . By the definition of  $\alpha$ -concavity,

$$F(z) \geq \prod_{j \in J} [F(v^j)]^{\lambda_j} \geq \prod_{j \in J} p^{\lambda_j} = p.$$

*Case 4:*  $\alpha \in (-\infty, 0)$ . By the definition of  $\alpha$ -concavity,

$$[F(z)]^\alpha \leq \sum_{j \in J} \lambda_j [F(v^j)]^\alpha \leq \sum_{j \in J} \lambda_j p^\alpha = p^\alpha.$$

Since  $\alpha < 0$ , we obtain  $F(z) \geq p$ .

*Case 5:*  $\alpha \in (0, \infty)$ . By the definition of  $\alpha$ -concavity,

$$[F(z)]^\alpha \geq \sum_{j \in J} \lambda_j [F(v^j)]^\alpha \geq \sum_{j \in J} \lambda_j p^\alpha = p^\alpha.$$

□

The consequence of this theorem is that under  $\alpha$ -concavity assumption all integer points contained in  $\text{co } \mathcal{Z} = E + \mathbb{R}_+^m$  satisfy the probabilistic constraint. This demonstrates the importance of the notion of  $\alpha$ -concave distribution function introduced in Definition 3. For example, the set  $\mathcal{Z}$  illustrated in Figure 2.1 does not correspond to any  $\alpha$ -concave distribution function, because its convex hull contains integer points which do not belong to  $\mathcal{Z}$ . These are the points (3,6), (4,5) and (6,2).

Under the conditions of Theorem 18, problem (2.8) can be formulated in the following equivalent way:

$$\begin{aligned} & \max f(x) \\ & \text{subject to } x \in \mathcal{D} \\ & g(x) \geq z, \end{aligned} \tag{2.9}$$

$$z \in \mathbb{Z}^m, \tag{2.10}$$

$$z \geq \sum_{j \in J} \lambda_j v^j \tag{2.11}$$

$$\sum_{j \in J} \lambda_j = 1$$

$$\lambda_j \geq 0, j \in J.$$

So, the probabilistic constraint has been replaced by algebraic equations and inequalities, together with the integrality requirement (2.10). This condition cannot be dropped, in general. However, if other conditions of the problem imply that  $g(x)$  is integer, we may dispose of  $z$  totally, and replace constraints (2.9)–(2.11) with

$$g(x) \geq \sum_{j \in J} \lambda_j v^j.$$

This may be the case for example, when we have an additional constraint in the definition of  $\mathcal{D}$  that  $x \in \mathbb{Z}^n$ , and  $g(x) = Tx$ , where  $T$  is a matrix of appropriate dimension with integer elements.

If  $Z$  takes values on a non-uniform grid, condition (2.10) should be replaced by the requirement that  $z$  is a grid point.

## 2.4 Optimality Conditions and Duality Theory

Let us split variables in problem (2.6):

$$\begin{aligned}
\max \quad & f(x) \\
& g(x) \geq z, \\
& x \in \mathcal{D}, \\
& z \in \mathcal{Z}.
\end{aligned} \tag{2.12}$$

We assume that  $p \in (0, 1)$ . Associating Lagrange multipliers  $u \in \mathbb{R}_+^m$  with constraints  $g(x) \geq z$ , we obtain the Lagrangian function

$$L(x, z, u) = f(x) + \langle u, g(x) - z \rangle.$$

The dual functional has the form

$$\Psi(u) = \sup_{(x, z) \in \mathcal{D} \times \mathcal{Z}} L(x, z, u) = h(u) - d(u),$$

where

$$h(u) = \sup\{f(x) + \langle u, g(x) \rangle \mid x \in \mathcal{D}\}, \tag{2.13}$$

$$d(u) = \inf\{\langle u, z \rangle \mid z \in \mathcal{Z}\}. \tag{2.14}$$

For any  $u \in \mathbb{R}_+^m$  the value of  $\Psi(u)$  is an upper bound on the optimal value  $F^*$  of the original problem. The best Lagrangian upper bound will be given by the optimal value  $D^*$  of the problem:

$$\inf_{u \geq 0} \Psi(u). \tag{2.15}$$

We call (2.15) the dual problem to problem (2.6). For  $u \not\geq 0$  one has  $d(u) = -\infty$ , because the set  $\mathcal{Z}$  contains a translation of  $\mathbb{R}_+^m$ . The function  $d(\cdot)$  is concave. Note that  $d(u) = -\sigma_{\mathcal{Z}}(-u)$ , where  $\sigma_{\mathcal{Z}}(\cdot)$  is the support function of the set  $\mathcal{Z}$ . By virtue of Theorem 15 and [161, Chapter V, Proposition 2.2.1], we have

$$d(u) = \inf\{\langle u, z \rangle \mid z \in \text{co } \mathcal{Z}\}. \tag{2.16}$$

Let us consider the *convex hull problem*:

$$\begin{aligned}
\max \quad & f(x) \\
& g(x) \geq z, \\
& x \in \mathcal{D}, \\
& z \in \text{co } \mathcal{Z}.
\end{aligned} \tag{2.17}$$

We make the following assumption.

**Constraint Qualification Condition.** *There exist points  $x^0 \in \mathcal{D}$  and  $z^0 \in \text{co } \mathcal{Z}$  such that  $g(x^0) > z^0$ .*

If the Constraint Qualification Condition is satisfied, from the duality theory in convex programming [298, Corollary 28.2.1] we know that there exists

$\hat{u} \geq 0$  at which the minimum in (2.15) is attained, and  $D^* = \Psi(\hat{u})$  is the optimal value of the convex hull problem (2.17).

We now study in detail the structure of the dual functional  $\Psi$ . We shall characterize the solution sets of the two subproblems (2.13) and (2.14), which provide values of the dual functional. Let us define the following sets:

$$\begin{aligned} V(u) &= \{v \in \mathbb{R}^m : \langle u, v \rangle = d(u) \text{ and } v \text{ is a } p\text{-efficient point}\}, \\ C(u) &= \{d \in \mathbb{R}_+^m : d_i = 0 \text{ if } u_i > 0, i = 1, \dots, m\}. \end{aligned} \quad (2.18)$$

**Lemma 5.** *For every  $u > 0$  the solution set of (2.14) is non-empty. For every  $u \geq 0$  it has the following form:  $\tilde{Z}(u) = V(u) + C(u)$ .*

**Proof.** Let us at first consider the case  $u > 0$ . Then every recession direction  $d$  of  $\mathcal{Z}$  satisfies  $\langle u, d \rangle > 0$ . Since  $\mathcal{Z}$  is closed, a solution to (2.14) must exist. Suppose that a solution  $z$  to (2.14) is not a  $p$ -efficient point. By virtue of Theorem 14, there is a  $p$ -efficient  $v \in \mathcal{Z}$  such that  $v \leq z$ , and  $v \neq z$ . Thus,  $\langle u, v \rangle < \langle u, z \rangle$ , which is a contradiction.

In the general case  $u \geq 0$ , the solution set of the problem to (2.14), if it is non-empty, always contains a  $p$ -efficient point. Indeed, if a solution  $z$  is not  $p$ -efficient, we must have a  $p$ -efficient point  $v$  dominated by  $z$ , and  $\langle u, v \rangle \leq \langle u, z \rangle$  holds by the non-negativity of  $u$ . Consequently,  $\langle u, v \rangle \leq \langle u, z \rangle$  for all  $p$ -efficient  $v \leq z$ , which is equivalent to  $z \in \{v\} + C(u)$ , as required.

If the solution set of (2.14) is empty then  $V(u) = \emptyset$  and the assertion is true as well.  $\square$

The last result allows us to calculate the subdifferential of  $d$  in a closed form.

**Lemma 6.** *For every  $u \geq 0$  one has  $\partial d(u) = \text{co } V(u) + C(u)$ . If  $u > 0$  then  $\partial d(u)$  is non-empty.*

**Proof.** From (2.14) we obtain  $d(u) = -\sigma_{\mathcal{Z}}(u)$ , where  $\sigma_{\mathcal{Z}}(\cdot)$  is the support function of  $\mathcal{Z}$  and, consequently, of  $\text{co } \mathcal{Z}$ . Recall that  $\sigma_{\mathcal{Z}}(u) = \delta_{\mathcal{Z}}^*(u)$ , where the latter is the conjugate of the indicator function of  $\mathcal{Z}$ . These facts follow from the structure of  $\mathcal{Z}$  described Theorem 14, by virtue of Corollary 16.5.1 in [298]. Thus

$$\partial d(u) = \partial \delta_{\mathcal{Z}}^*(-u).$$

Recall that  $\text{co } \mathcal{Z}$  is closed, by Theorem 15. Using [298, Theorem 23.5], we observe that  $s \in \partial \delta_{\mathcal{Z}}^*(-u)$  if and only if  $\delta_{\mathcal{Z}}^*(-u) + \delta_{\text{co } \mathcal{Z}}(s) = -\langle s, u \rangle$ , where  $\delta_{\text{co } \mathcal{Z}}(\cdot)$  is the indicator function of  $\text{co } \mathcal{Z}$ . It follows that  $s \in \text{co } \mathcal{Z}$  and  $\delta_{\mathcal{Z}}^*(-u) = -\langle s, u \rangle$ . Consequently,

$$\langle s, u \rangle = d(u). \quad (2.19)$$

Since  $s \in \text{co } \mathcal{Z}$  we can represent it as follows:

$$s = \sum_{j=1}^{m+1} \alpha_j e^j + w,$$

where  $e^j$ ,  $j = 1, \dots, m+1$ , are extreme points of  $\text{co } \mathcal{Z}$  and  $w \geq 0$ . Using Theorem 16 we conclude that  $e^j$  are  $p$ -efficient points. Moreover

$$\langle s, u \rangle = \sum_{j=1}^{m+1} \alpha_j \langle u, e^j \rangle + \langle u, w \rangle \geq d(u), \quad (2.20)$$

because  $\langle u, e^j \rangle \geq d(u)$  and  $\langle u, w \rangle \geq 0$ . Combining (2.19) and (2.20) we conclude that  $\langle u, e^j \rangle = d(u)$  for all  $j$ , and  $\langle u, w \rangle = 0$ . Thus  $s \in \text{co } V(u) + C(u)$ .

Conversely, if  $s \in \text{co } V(u) + C(u)$  then (2.19) holds true. This implies that  $s \in \partial d(u)$ , as required.

The set  $\partial d(u)$  is non-empty for  $u > 0$  by virtue of Lemma 5.  $\square$

Now we analyze the function  $h(\cdot)$ . Define the set of maximizers in (2.13),

$$X(u) = \{x \in \mathcal{D} : f(x) + \langle u, g(x) \rangle = h(u)\}.$$

By the convexity of the set  $\mathcal{D}$  and by the concavity of  $f$  and  $g$ , the solution set  $X(u)$  is convex for all  $u \geq 0$ .

**Lemma 7.** *Assume that the set  $\mathcal{D}$  is compact. The subdifferential of the function  $h$  is described as follows for every  $u \in \mathbb{R}^m$ :*

$$\partial h(u) = \text{co} \{g(x) : x \in X(u)\}.$$

**Proof.** The function  $h$  is convex on  $\mathbb{R}^m$ . Since the set  $\mathcal{D}$  is compact and  $f$  and  $g$  are concave, the set  $X(u)$  is compact. Therefore, the subdifferential of  $h(u)$  for every  $u \in \mathbb{R}^m$  is the closure of  $\text{co} \{g(x) : x \in X(u)\}$  (see [161, Chapter VI, Lemma 4.4.2]). By the compactness of  $X(u)$  and concavity of  $g$ , the set  $\{g(x) : x \in X(u)\}$  is closed. Therefore, we can omit taking the closure in the description of the subdifferential of  $h(u)$ .  $\square$

This analysis provides the basis for the following necessary and sufficient optimality conditions for problem (2.15).

**Theorem 19.** *Assume that the Constraint Qualification Condition is satisfied and that the set  $\mathcal{D}$  is compact. A vector  $u \geq 0$  is an optimal solution of (2.15) if and only if there exists a point  $x \in X(u)$ , points  $v^1, \dots, v^{m+1} \in V(u)$  and scalars  $\beta_1, \dots, \beta_{m+1} \geq 0$  with  $\sum_{j=1}^{m+1} \beta_j = 1$ , such that*

$$g(x) - \sum_{j=1}^{m+1} \beta_j v^j \in C(u), \quad (2.21)$$

where  $C(u)$  is given by (2.18).

**Proof.** Since  $-C(u)$  is the normal cone to the positive orthant at  $u \geq 0$ , the necessary and sufficient optimality condition for (2.15) has the form

$$\partial \Psi(u) \cap C(u) \neq \emptyset \quad (2.22)$$



(cf. [298, Theorem 27.4]). Since  $\text{int dom } d \neq \emptyset$  and  $\text{dom } h = \mathbb{R}^m$  we have  $\partial\Psi(u) = \partial h(u) - \partial d(u)$ . Using Lemma 6 and Lemma 7, we conclude that there exist

$$\begin{aligned} & p\text{-efficient points } v^j \in V(u), \quad j = 1, \dots, m+1, \\ & \beta^j \geq 0, \quad j = 1, \dots, m+1, \quad \sum_{j=1}^{m+1} \beta_j = 1, \\ & x^j \in X(u), \quad j = 1, \dots, m+1, \\ & \alpha^j \geq 0, \quad j = 1, \dots, m+1, \quad \sum_{j=1}^{m+1} \alpha_j = 1, \end{aligned}$$

such that

$$\sum_{j=1}^{m+1} \alpha_j g(x^j) - \sum_{j=1}^{m+1} \beta_j v^j \in C(u). \quad (2.23)$$

If the functions  $f$  and  $g$  were strictly concave, the set  $X(u)$  would be a singleton. Then all  $x^j$  would be identical and the above relation would immediately imply (2.21). Otherwise, let us define

$$x = \sum_{j=1}^{m+1} \alpha_j x^j.$$

By the convexity of  $X(u)$  we have  $x \in X(u)$ . Consequently,

$$f(x) + \sum_{i=1}^m u_i g_i(x) = h(u) = f(x^j) + \sum_{i=1}^m u_i g_i(x^j), \quad j = 1, \dots, m+1.$$

Multiplying the last equation by  $\alpha_j$  and adding we obtain

$$f(x) + \sum_{i=1}^m u_i g_i(x) = \sum_{j=1}^{m+1} \alpha_j \left[ f(x^j) + \sum_{i=1}^m u_i g_i(x^j) \right].$$

Since  $g_i(x) \geq \sum_{j=1}^{m+1} \alpha_j g_i(x^j)$ , substituting into the above equation, we obtain

$$f(x) \leq \sum_{j=1}^{m+1} \alpha_j f(x^j).$$

If  $g_i(x) > \sum_{j=1}^{m+1} \alpha_j g_i(x^j)$  and  $u_i > 0$  for some  $i$ , the above inequality becomes strict, in contradiction to the concavity of  $f$ . Thus, for all  $u_i > 0$  we have  $g_i(x) = \sum_{j=1}^{m+1} \alpha_j g_i(x^j)$ , and it follows that

$$g(x) - \sum_{j=1}^{m+1} \alpha_j g(x^j) \in C(u).$$

Since  $C(u)$  is a convex cone, we can combine the last relation with (2.23) and obtain (2.21), as required.

Now we prove the converse implication. Assume that we have  $x \in X(u)$ , points  $v^1, \dots, v^{m+1} \in V(u)$  and scalars  $\beta_1, \dots, \beta_{m+1} \geq 0$  with  $\sum_{j=1}^{m+1} \beta_j = 1$ , such that (2.21) holds true. By Lemma 6 and Lemma 7 we have

$$g(x) - \sum_{j=1}^{m+1} \beta_j v^j \in \partial \Psi(u).$$

Thus (2.21) implies (2.22), which is a necessary and sufficient optimality condition for (2.15).  $\square$

Since the set of  $p$ -efficient points is not known, we need a numerical method for solving the convex hull problem (2.17) or its dual (2.15).

Using these optimality conditions we obtain the following duality result.

**Theorem 20.** *Assume that the Constraint Qualification Condition for problem (2.12) is satisfied, the probability distribution of the vector  $Z$  is  $\alpha$ -concave for some  $\alpha \in [-\infty, \infty]$ , and the set  $\mathcal{D}$  is compact. If a point  $(\hat{x}, \hat{z})$  is an optimal solution of (2.12), then there exists a vector  $\hat{u} \geq 0$ , which is an optimal solution of (2.15) and the optimal values of both problems are equal. If  $\hat{u}$  is an optimal solution of problem (2.15), then there exist a point  $\hat{x}$ , such that  $(\hat{x}, g(\hat{x}))$  is a solution of problem (2.12), and the optimal values of both problems are equal.*

**Proof.** From the  $\alpha$ -concavity assumption we obtain that problems (2.12) and (2.17) coincide. If  $\hat{u}$  is optimal solution of problem (2.15), we obtain the existence of points  $\hat{x} \in X(\hat{u})$ ,  $v^1, \dots, v^{m+1} \in V(u)$  and scalars  $\beta_1, \dots, \beta_{m+1} \geq 0$  with  $\sum_{j=1}^{m+1} \beta_j = 1$ , such that the optimality conditions in Theorem 19 are satisfied. Setting  $\hat{z} = g(\hat{x})$  we have to show that  $(\hat{x}, \hat{z})$  is an optimal solution of problem (2.12) and that the optimal values of both problems are equal. First we observe that this point is feasible. Set  $s \in C(\hat{u}) : s = g(\hat{x}) - \sum_{j=1}^{m+1} \beta_j v^j$ . From the definitions of  $X(\hat{u})$ ,  $V(\hat{u})$ , and  $C(\hat{u})$  we obtain

$$\begin{aligned} h(\hat{u}) &= f(\hat{x}) + \langle \hat{u}, g(\hat{x}) \rangle = f(\hat{x}) + \langle \hat{u}, \sum_{j=1}^{m+1} \beta_j v^j + s \rangle \\ &= f(\hat{x}) + \sum_{j=1}^{m+1} \beta_j d(\hat{u}) + \langle \hat{u}, s \rangle = f(\hat{x}) + d(\hat{u}). \end{aligned}$$

Thus,  $f(\hat{x}) = h(\hat{u}) - d(\hat{u}) = D^* \geq F^*$ , which proves that  $(\hat{x}, \hat{z})$  is an optimal solution of problem (2.12) and  $D^* = F^*$ .

If  $(\hat{x}, \hat{z})$  is a solution of (2.12), then by [298, Theorem 28.4] there is a vector  $\hat{u} \geq 0$  such that  $\hat{u}_i(\hat{z}_i - g_i(\hat{x})) = 0$  and  $\partial f(\hat{x}) + \partial \langle \hat{u}, g(\hat{x}) - \hat{z} \rangle \cap -\partial [\delta_{\mathcal{D}}(\hat{x}) + \delta_{\mathcal{Z}}(\hat{z})] \neq \emptyset$ , where  $\delta_C(\cdot)$  denotes the indicator function of the set  $C$ . Thus, there are vectors

$$s \in \partial f(\hat{x}) + \partial \langle u, g(\hat{x}) \rangle \cap -\partial \delta_{\mathcal{D}}(\hat{x}) \quad (2.24)$$

and

$$\hat{u} \in \partial \delta_{\mathcal{Z}}(\hat{z}). \quad (2.25)$$

The first inclusion (2.24) is the optimality condition for problem (2.13), and thus  $x \in X(\hat{u})$ . By virtue of [298, Theorem 23.5] the inclusion (2.25) is equivalent to  $\hat{z} \in \partial \delta_{\mathcal{Z}}^*(\hat{u})$ . Using Lemma 6 we obtain that  $\hat{z} \in \partial d(\hat{u}) = \text{co}V(\hat{u}) + C(\hat{u})$ . Thus, there exists points  $v^1, \dots, v^{m+1} \in V(u)$  and scalars  $\beta_1, \dots, \beta_{m+1} \geq 0$  with  $\sum_{j=1}^{m+1} \beta_j = 1$ , such that  $\hat{z} - \sum_{j=1}^{m+1} \beta_j v^j \in C(\hat{u})$ . Using the complementarity condition  $\hat{u}_i(\hat{z}_i - g_i(\hat{x})) = 0$  we conclude that the optimality conditions of Theorem 19 are satisfied. Thus  $\hat{u}$  is an optimal solution of (2.15).  $\square$

For the special case of discrete distribution and linear constraints we can obtain a more specific necessary and sufficient condition for the existence of an optimal solution of (2.8).

The *linear* probabilistic optimization problem assumes that  $g(x) = Tx$ , where  $T$  is an  $m \times n$  matrix,  $f(x) = \langle c, x \rangle$  with  $c \in \mathbb{R}^n$ . Furthermore,  $\mathcal{D}$  is a convex closed polyhedron in  $\mathbb{R}^n$ . It is usually formulated as follows:

$$\begin{aligned} & \min \langle c, x \rangle \\ & \text{subject to } \mathbb{P}[Tx \geq Z] \geq p \\ & \quad Ax \geq b \\ & \quad x \geq 0. \end{aligned} \quad (2.26)$$

Here  $A$  is an  $s \times n$  matrix and  $b \in \mathbb{R}^s$ .

**Assumption 2.1.** *The set  $\Lambda := \{(u, w) \in \mathbb{R}_+^{m+s} \mid A^T w + T^T u \leq c\}$  is non-empty.*

**Theorem 21.** *Assume that the feasible set of (2.26) is non-empty and that  $Z$  has a discrete distribution. Then (2.26) has an optimal solution if and only if Assumption 2.1 holds.*

**Proof.** If (2.8) has an optimal solution, then for some  $j \in J$  the linear optimization problem

$$\begin{aligned} & \min \langle c, x \rangle \\ & \text{subject to } Tx \geq v^j \\ & \quad Ax \geq b \\ & \quad x \geq 0 \end{aligned} \quad (2.27)$$

has an optimal solution. By duality in linear programming, its dual problem

$$\begin{aligned}
& \max \langle v^j, u \rangle + \langle b, w \rangle \\
& \text{subject to } T^T u + A^T w \leq c \\
& \quad u, w \geq 0
\end{aligned} \tag{2.28}$$

has an optimal solution and the optimal values of both programs are equal. Thus, Assumption 2.1 must hold. On the other hand, if Assumption 2.1 is satisfied, all dual programs (2.28) for  $j \in J$  have non-empty feasible sets, so the objective values of all primal problems (2.27) are bounded from below. Since one of them has a non-empty feasible set by assumption, an optimal solution must exist.  $\square$

## 2.5 Methods for Solving Probabilistic Programming Problems

When the constraint functions  $g_i(x, Z)$ ,  $i = 1, \dots, m$  are not separable the optimization problem is difficult to handle. Numerical methods for these problems are based on combinatorial techniques (see Section 2.5.8), on response surface approximations (see Section 2.5.6), or via Monte Carlo methods (see Chapter 5 in this volume).

Numerical techniques for probabilistic problems with random right hand sides ( $g(x, Z) := \tilde{g}(x) - Z$ ) are much better developed, particularly for linear function  $\tilde{g}(x) = Tx$ . If the distribution of  $Z$  is  $\alpha$ -concave, it follows from Corollary 2 that the feasible set of this problem is convex. Therefore, methods of convex programming can be applied. The specificity here is in the implicit definition of the feasible set and in the difficulty to evaluate the constraint function

$$G(x) = \mathbb{P}[Tx \geq Z] = F_Z(Tx).$$

It is even more difficult to estimate its gradient if it exists. Specialized Monte Carlo integration techniques have been developed for some classes of distributions of  $Z$ , in particular for the normal distribution and for the gamma distribution (see [223, 224, 349–351]).

We review some of the known methods and present in more detail two recent methods for solving nonlinear probabilistic problems: the dual method in Section 2.5.3 and the primal-dual method in Section 2.5.4. Both of these methods are based on the duality analysis presented in the previous section.

### 2.5.1 A Cutting Plane Method

One of the first methods for probabilistic optimization is based on cutting planes techniques for the following problem:

$$\begin{aligned}
& \min f(x) \\
& \text{subject to } \mathbb{P}[Tx \geq Z] \geq p \\
& \quad Ax = b, x \geq 0.
\end{aligned} \tag{2.29}$$

It is assumed that the constraint function  $G(x) = \mathbb{P}[Tx \geq Z]$  is quasi-concave and it has continuous gradients. Additionally, we assume that there exists a bounded convex polyhedron  $C^1$  containing the set of feasible solutions of problem (2.29).

Furthermore, the following constraint qualification condition is satisfied: there exists a vector  $x^0$  such that

$$G(x^0) > p, \quad x^0 \in \{x \mid Ax = b, x \geq 0\} \quad (2.30)$$

The algorithm works in two phases. In the first phase a feasible point  $x^0$  satisfying the constraint qualification condition is found. This can be done by maximizing  $G(x)$  subject to the constraints  $Ax = b, x \geq 0$ , by using any gradient method. Of course in our case, we do not need to carry out all steps in the gradient descent method, it is sufficient to find a point  $x^*$  such that  $G(x^*) > p$ .

The second phase consists of the following steps.

Step 1. Solve the problem:

$$\begin{aligned} & \min f(x) \\ & \text{subject to } x \in C^k. \end{aligned}$$

Let  $x^k$  be an optimal solution. If  $x^k$  is feasible, then stop,  $x^k$  is an optimal solution of problem (2.29).

Step 2. Let  $\lambda^k$  be the largest  $\lambda \geq 0$  such that  $x^0 + \lambda(x^k - x^0)$  is feasible and set

$$y^k = x^0 + \lambda^k(x^k - x^0).$$

If  $G(y^k) = p$ , then define

$$C^{k+1} = \{x \mid x \in C^k, \nabla G(y^k)(x - y^k) \geq 0\}.$$

Consider any other constraint that is active at  $y^k$  and set  $C^{k+1}$  to be the intersection of  $C^k$  and the set determined by this constraint. Go to Step 1.

### 2.5.2 The Logarithmic Barrier Function Method

If  $Z$  has a log-concave distribution on  $\mathbb{R}^n$  then the constraint function of problem (2.29) is log-concave on the set  $\{x \in \mathbb{R}^n : G(x) \geq p\}$ . We can solve problem (2.29) by using logarithmic penalty functions. We take a decreasing sequence of positive numbers  $\{s^k\}$  such that  $\lim_{k \rightarrow \infty} s^k = 0$  and solve the problem

$$\begin{aligned} & \min \{f(x) - s^k \log(G(x) - p)\} \\ & \text{subject to } Ax = b \\ & \quad x \geq 0. \end{aligned}$$

We obtain a point  $x^k$  as an optimal solution of this problem. The sequence  $\{f(x^k)\}$  converges to the optimal value of problem (2.29) under the assumptions that  $f(\cdot)$  is a continuous function,  $G(\cdot)$  is continuous and log-concave, constraint qualification condition (2.30) is satisfied, and the set  $\{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$  is bounded.

A modern primal-dual interior point method using similar idea is developed in [329].

### 2.5.3 The Dual Method

This method has been proposed in [104] for solving nonlinear probabilistic problems of form (2.6). The idea of the method is to solve the dual problem (2.15) using the information about the subgradients of the dual functional  $\Psi$  to generate convex piecewise-linear approximations of  $\Psi$ . Suppose that the values of the functional  $\Psi$  at certain points  $u^j$ ,  $j = 1, \dots, k$ , are available. Moreover, we assume that the corresponding solutions  $v^1, \dots, v^k$  and  $x^1, \dots, x^k$  of the two problems (2.16) and (2.13) are available as well. According to Lemma 5 we can assume that  $v^j$ ,  $j = 1, \dots, k$  are  $p$ -efficient points. By virtue of Lemma 7 and Lemma 6 the following function  $\Psi_k(\cdot)$  is a lower bound of  $\Psi$ :

$$\Psi_k(u) := \max_{1 \leq j \leq k} [\Psi(u^j) + \langle g(x^j) - v^j, u - u^j \rangle].$$

Minimizing  $\Psi_k(u)$  over  $u \geq 0$ , we obtain the next iterate  $u^{k+1}$ . For the purpose of numerical tractability, we shall impose an upper bound  $b \in \mathbb{R}$  on the dual variables  $u_j$ . We define the feasible set of the dual problem as follows:

$$U := \{u \in \mathbb{R}^m : 0 \leq u_i \leq b, i = 1, \dots, m\}$$

where  $b$  is a sufficiently large number. We also use  $\varepsilon > 0$  as a stopping test parameter.

The algorithm works as follows:

Step 0. Select a vector  $u^1 \in U$ . Set  $\Psi_0(u^1) = -\infty$  and  $k = 1$ .

Step 1. Calculate

$$h(u^k) = \max\{f(x) + \langle u^k, g(x) \rangle \mid x \in \mathcal{D}\}, \quad (2.31)$$

$$d(u^k) = \min\{\langle u^k, z \rangle \mid z \in \text{co } \mathcal{Z}\}. \quad (2.32)$$

Let  $x^k$  be the solution of problem (2.31) and  $v^k$  be the solution of problem (2.32).

Step 2. Calculate  $\Psi(u^k) = h(u^k) - d(u^k)$ . If  $\Psi(u^k) \leq \Psi_{k-1}(u^k) + \varepsilon$  then stop; otherwise continue.

Step 3. Define

$$\Psi_k(u) = \max_{1 \leq j \leq k} [\Psi(u^j) + \langle g(x^j) - v^j, u - u^j \rangle],$$

and find a solution  $u^{k+1}$  of the problem

$$\min_{u \in U} \Psi_k(u).$$

Step 4. Increase  $k$  by one and go to Step 1.

Problem (2.31) is a convex nonlinear problem, and it can be solved by a suitable numerical method for nonlinear optimization. Problem (2.32) requires a dedicated numerical method. In particular applications, specialized methods may provide its efficient numerical solution. Alternatively, one can approximate the random vector  $Y$  by finitely many realizations (scenarios). More detailed discussion on this idea will follow in Section 2.5.7.

**Theorem 22.** *Suppose that  $\varepsilon = 0$ . Then the sequences  $\Psi(u^k)$  and  $\Psi_k(u^k)$ ,  $k = 1, 2, \dots$ , converge to the optimal value of problem (2.15). Moreover, every accumulation point of the sequence  $\{u^k\}$  is an optimal solution of (2.15).*

**Proof.** The convergence of the method follows from a standard argument about cutting plane methods for convex optimization (see, e.g., [161, Theorem 4.2.3]).  $\square$

Let us discuss a way of recovering a primal solution from the sequences of points  $\{u^k\}$ ,  $\{x^k\}$  and  $\{v^k\}$  generated by the method.

It follows from Theorem 22 that for every  $\varepsilon > 0$  the dual method has to stop after finitely many iterations at some step  $k$  for which

$$\Psi(u^k) - \varepsilon \leq \Psi_{k-1}(u^k) \leq \min_{u \in U} \Psi(u). \quad (2.33)$$

Let us define the set of active cutting planes at  $u^k$ :

$$J = \{j \in \{1, \dots, k-1\} : \Psi(u^j) + \langle g(x^j) - v^j, u^k - u^j \rangle = \Psi_{k-1}(u^k)\}.$$

The subdifferential of  $\Psi_{k-1}(\cdot)$  has the form

$$\partial \Psi_{k-1}(u) = \left\{ s \in \mathbb{R}^m : s = \sum_{j \in J} \alpha_j (g(x^j) - v^j), \sum_{j \in J} \alpha_j = 1, \alpha_j \geq 0, j \in J \right\}.$$

Since  $u^k$  is a minimizer of  $\Psi_{k-1}(\cdot)$ , there must exist a subgradient  $s$  such that

$$s \in C(u^k).$$

Thus there exist non-negative  $\alpha_j$  totaling 1 such that

$$\sum_{j \in J} \alpha_j (g(x^j) - v^j) \in C(u^k). \quad (2.34)$$

By the definition of  $\Psi$ ,

$$\Psi(u^j) = f(x^j) + \langle u^j, g(x^j) \rangle - \langle u^j, v^j \rangle.$$

Substituting this into the definition of the set  $J$  we obtain that

$$\Psi_{k-1}(u^k) = f(x^j) + \langle g(x^j) - v^j, u^k \rangle, \quad j \in J.$$

Multiplying both sides by  $\alpha_j$  and summing up we conclude that

$$\Psi_{k-1}(u^k) = \sum_{j \in J} \alpha_j f(x^j) + \left\langle \sum_{j \in J} \alpha_j (g(x^j) - v^j), u^k \right\rangle.$$

This combined with (2.34) yields

$$\Psi_{k-1}(u^k) = \sum_{j \in J} \alpha_j f(x^j). \quad (2.35)$$

Define

$$\bar{x} = \sum_{j \in J} \alpha_j x^j, \quad \bar{z} = \sum_{j \in J} \alpha_j v^j.$$

Clearly,  $\bar{x} \in \mathcal{D} \cap \text{co } \mathcal{Z}$ . Using the concavity of  $g$  and (2.34) we see that

$$g(\bar{x}) \geq \sum_{j \in J} \alpha_j g(x^j) \geq \sum_{j \in J} \alpha_j v^j = \bar{z}.$$

Thus the point  $(\bar{x}, \bar{z})$  is feasible for the convex hull problem (2.17).

It follows from the concavity of  $f$  and (2.35) that

$$f(\bar{x}) \geq \sum_{j \in J} \alpha_j f(x^j) = \Psi_{k-1}(u^k).$$

By the stopping test (2.33),

$$f(\bar{x}) \geq \Psi(u^k) - \varepsilon. \quad (2.36)$$

Since the value of  $\Psi(u)$  is an upper bound for the objective value at any feasible point  $(x, z)$  of the convex hull problem, we conclude that  $(\bar{x}, \bar{z})$  is an  $\epsilon$ -optimal solution of this problem.

The above construction can be carried out at every iteration  $k$ . In this way we obtain a certain sequence  $(\bar{x}^k, \bar{v}^k)$ ,  $k = 1, 2, \dots$ . Since the sequence  $\{\bar{x}^k\}$  is contained in a compact set and each  $(\bar{x}^k, \bar{z}^k)$  is feasible for the convex hull problem (2.17), the sequence  $\{\bar{z}^k\}$  is included in a compact set as well. Thus the sequence  $\{(\bar{x}^k, \bar{v}^k)\}$  has accumulation points. It follows from Theorem 22 and from (2.36) that every accumulation point of the sequence  $\{(\bar{x}^k, \bar{v}^k)\}$  is a solution of the convex hull problem (2.17). Under the assumptions of Corollary 2 the accumulation point is a solution of problem (2.6).



### 2.5.4 The Primal-Dual Method

This approach was first suggested for linear probabilistic problems in [191]. Involving tools of non-smooth analysis the method was successfully developed for nonlinear probabilistic optimization and general distributions in [104].

The algorithm presented in the previous section is based on a cutting plane approximation of the entire dual functional. The method of this section involves approximations of the functional  $d(\cdot)$  only. The method consists of an iterative generation of  $p$ -efficient points and the solution of a restriction of problem (2.4). The restriction is based on the disjunctive representation of  $\text{co } \mathcal{Z}$  by the  $p$ -efficient points generated so far.

We assume that we know a compact set  $B$  containing all  $p$ -efficient points  $v$  such that there exists  $x \in \mathcal{D}$  satisfying  $v \leq g(x)$ . It may be just a box with the lower bound at  $l$ , the vector of  $p$ -efficient points of all marginal distributions of  $Y$ , and with the upper bound above the maxima of  $g_i(x)$  over  $x \in \mathcal{D}$ ,  $i = 1, \dots, m$ . Such a box exists by the compactness of  $\mathcal{D}$ . We also use a stopping test parameter  $\varepsilon > 0$ .

We denote the simplex in  $\mathbb{R}^k$  by  $S_k$ , *i.e.*,

$$S_k := \{\lambda \in \mathbb{R}^k : \lambda_i \geq 0, \sum_{i=1}^k \lambda_i = 1\}.$$

The primal-dual method follows the steps:

Step 0. Select a  $p$ -efficient point  $v^1 \in B$  such that there exists  $\tilde{x} \in \mathcal{D}$  satisfying  $g(\tilde{x}) > v^1$ . Set  $J_1 = \{1\}$ ,  $k = 1$ .

Step 1. Solve the *master problem*

$$\max f(x) \tag{2.37}$$

$$g(x) \geq \sum_{j \in J_k} \lambda_j v^j, \tag{2.38}$$

$$x \in \mathcal{D}, \lambda \in S_k. \tag{2.39}$$

Let  $u^k$  be the vector of Lagrange multipliers associated with the constraint (2.38).

Step 2. Calculate  $d_k(u^k) = \min_{j \in J_k} \langle u^k, v^j \rangle$ .

Step 3. Find a  $p$ -efficient solution  $v^{k+1}$  of the subproblem:

$$\min_{z \in \mathcal{Z} \cap B} \langle u^k, z \rangle$$

and calculate  $d(u^k) = \langle v^{k+1}, u^k \rangle$ .

Step 4. If  $d(u^k) \geq d_k(u^k) - \varepsilon$  then stop; otherwise set  $J_{k+1} = J_k \cup \{k+1\}$ , increase  $k$  by one, and go to Step 1.

The first  $p$ -efficient point  $v^1$  can be found by solving the subproblem at Step 3 for some  $u \geq 0$ . All master problems will be solvable, if the first

one is solvable, which is assumed at Step 0. Moreover, all master problems satisfy Slater's constraint qualification condition with the point  $\tilde{x}$  and  $\tilde{\lambda} = (1, 0, \dots, 0)$ . Therefore, it is legitimate to assume at Step 1 that we obtain a vector of Lagrange multipliers associated with (2.38). The subproblem at Step 3 is the same as (2.32) in the dual method. It requires a dedicated approach.

**Theorem 23.** *Let  $\varepsilon = 0$ . The sequence  $\{f(x^k)\}$ ,  $k = 1, 2, \dots$  converges to the optimal value of the convex hull problem (2.17). Every accumulation point  $\hat{x}$  of the sequence  $\{x^k\}$  is an optimal solution of problem (2.17), with  $z = g(\hat{x})$ .*

**Proof.** We formulate the dual problem to the master problem (2.37)–(2.39).

The dual functional is defined as follows:

$$\Phi_k(u) = \sup \left\{ f(x) + \langle u, g(x) - \sum_{j \in J_k} \lambda_j v^j \rangle : x \in \mathcal{D}, \lambda \in S_k \right\} = h(u) - d_k(u),$$

where  $h(u)$  is the same as in (2.13) and

$$d_k(u) = \inf_{\lambda \in S_k} \sum_{j \in J_k} \lambda_j \langle u, v^j \rangle.$$

It is clear that  $d_k(u) = \min_{j \in J_k} \langle u, v^j \rangle \geq d(u)$ , where  $d(u)$  is as in (2.14). Thus the function  $\Phi_k(u)$  is a lower bound of the dual functional  $\Psi(u)$ , i.e.,

$$\Phi_k(u^k) \leq \Psi(u^k).$$

Since  $J_k \subset J_{k+1}$ , for every feasible point  $(x, \lambda)$  of problem (2.37)–(2.39) at iteration  $k$ , the point  $(x, (\lambda, 0))$  is feasible at iteration  $k+1$ . Therefore the sequence  $\{f(x^k)\}$  is monotonically increasing. By duality, the sequence  $\{\Phi_k(u^k)\}$  is monotonically increasing as well.

For  $\delta > 0$  consider the set  $K_\delta$  of iteration numbers  $k$  for which

$$\Phi_k(u^k) + \delta \leq \Psi(u^k).$$

Suppose that  $k \in K_\delta$ . We obtain the following chain of inequalities for all  $j \leq k$ :

$$\begin{aligned} \delta &\leq \Psi(u^k) - \Phi_k(u^k) = -d(u^k) + d_k(u^k) = -\min_{z \in \mathcal{Z} \cap B} \langle u^k, z \rangle + \min_{j \in J_k} \langle u^k, v^j \rangle \\ &\leq \langle u^k, v^j - v^{k+1} \rangle \leq \|u^k\| \cdot \|v^j - v^{k+1}\|. \end{aligned}$$

We shall show later that there exists  $M > 0$  such that  $\|u^k\| \leq M$  for all  $k$ . Therefore

$$\|v^{k+1} - v^j\| \geq \delta/M \quad \text{for all } k \in K_\delta \quad \text{and all } j = 1, \dots, k.$$

It follows from the compactness of the set  $B$  that the set  $K_\delta$  is finite for every  $\delta > 0$ . Thus, we can find a subsequence  $\mathcal{K}$  such that

$$\Psi(u^k) - \Phi_k(u^k) \rightarrow 0, \quad k \in \mathcal{K}.$$

Since for all  $k$

$$\Psi(u^k) \geq \min_{u \geq 0} \Psi(u) \geq \min_{u \geq 0} \Phi_k(u) = \Phi_k(u^k), \quad (2.40)$$

and the sequence  $\{\Phi_k(u^k)\}$  is nondecreasing, we conclude that

$$\lim_{k \rightarrow \infty} \Phi_k(u^k) = \min_{u \geq 0} \Psi(u).$$

We also have  $\Phi_k(u^k) = f(x^k)$  and thus the sequence  $\{f(x^k)\}$  is convergent to the optimal value of the convex hull problem (2.17). Since  $\{x^k\}$  is included in  $\mathcal{D}$ , it has accumulation points and every accumulation point  $\hat{x}$  is a solution of (2.17), with  $z = g(\hat{x})$ .

It remains to show that the multipliers  $u^k$  are uniformly bounded. To this end observe that the Lagrangian

$$L_k(x, \lambda, u^k) = f(x) + \langle u^k, g(x) - \sum_{j=1}^k \lambda_j v^j \rangle$$

achieves its maximum in  $\mathcal{D} \times S_k$  at  $x^k$  and some  $\lambda^k$ . The optimal value is equal to  $f(x^k)$  and it is bounded above by the optimal value  $\mu$  of the convex hull problem (2.17).

The point  $\tilde{x}$  and  $\tilde{\lambda} = (1, 0, \dots, 0)$  is in  $\mathcal{D} \times S_k$ . Therefore

$$L_k(\tilde{x}, \tilde{\lambda}, u^k) \leq \mu.$$

It follows that

$$\langle u^k, g(\tilde{x}) - v^1 \rangle \leq \mu - f(\tilde{x}).$$

Recall that  $g(\tilde{x}) - v^1 > 0$ . Therefore  $u^k$  is an element of the compact set

$$U = \{u \in \mathbb{R}^m : \langle u, g(\tilde{x}) - v^1 \rangle \leq \mu - f(\tilde{x}), u \geq 0\}.$$

If we use  $\varepsilon > 0$  at Step 4, then relations (2.40) guarantee that the current solution  $x^k$  is  $\varepsilon$ -optimal for the convex hull problem (2.17). □

Under the assumption that the distribution function of the random vector  $Y$  is  $\alpha$ -concave for some  $\alpha \in \overline{\mathbb{R}}$ , the suggested algorithms provide an optimal solution of problem (2.4). Otherwise, we obtain an upper bound of the optimal value. Moreover, the solution point  $\hat{x}$  determined by both algorithms satisfies the constraint  $g(x) \in \text{co } \mathcal{Z}$ , and may not satisfy the probabilistic constraint.

We now suggest an approach to determine a primal feasible solution.

Both the dual and the primal-dual method end with a collection of  $p$ -efficient points. In the primal-dual algorithm, we consider the multipliers  $\lambda_j$  of the master problem (2.37)–(2.39). We define  $C = \{j \in J : \lambda_j > 0\}$ . In the dual algorithm, we consider the active cutting planes in the last approximation, and

set  $C = \{j \in J : \beta_j > 0\}$ , where  $J$  and  $\beta_j$  are determined in the proof of Theorem 22.

In both cases, if  $C$  contains only one element, the point  $\hat{x}$  is feasible and therefore optimal for the disjunctive formulation (2.7). If, however, there are more elements in  $C$ , we need to find a feasible point. A natural possibility is to consider the *restricted* disjunctive formulation:

$$\begin{aligned} & \max f(x) \\ & \text{subject to } g(x) \in \bigcup_{j \in C} K_j, \\ & x \in \mathcal{D}. \end{aligned} \tag{2.41}$$

It can be solved by simple enumeration of all cases for  $j \in C$ :

$$\begin{aligned} & \max f(x) \\ & \text{subject to } g(x) \geq v^j, \\ & x \in \mathcal{D}. \end{aligned} \tag{2.42}$$

An alternative strategy would be to solve the corresponding bounding problem (2.42) every time a new  $p$ -efficient point is generated. If  $\mu_j$  denotes the optimal value of (2.42), the lower bound at iteration  $k$  is

$$\bar{\mu}^k = \max_{0 \leq j \leq k} \mu_j.$$

A quantitative estimate of the errors and the comparison of both methods are difficult and require new techniques.

### 2.5.5 Nonparametric Estimates of Distribution Functions

In this subsection we shall assume that the probabilistic constraint is formulated as follows:

$$\mathbb{P}(Tx \leq Z) \geq p.$$

Furthermore, we assume that the random variables  $Z_1, \dots, Z_m$  are independent and each has a continuous distribution with density  $h_i(\cdot)$ . Using the marginal distribution functions  $F_i(z) = \mathbb{P}(Z_i \leq z)$ , problem (2.26) can be written in the following equivalent form:

$$\begin{aligned} & \min \langle c, x \rangle \\ & \text{subject to } \prod_{i=1}^m (1 - F_i(z_i)) \geq p \\ & T_i x = z_i, \quad i = 1, \dots, m \\ & Ax = b \\ & x \geq 0. \end{aligned} \tag{2.43}$$

If for any feasible solution  $x$  of the this problem the probabilistic constraint is satisfied as a strict inequality, we can take logarithm on both sides of this constraint.

We define the auxiliary functions:

$$g_i(t) = \begin{cases} \frac{h_i(t)}{1-F_i(t)} & \text{if } F_i(t) < 1 \\ 0 & \text{if } F_i(t) = 1 \end{cases} \quad (2.44)$$

Assuming that the functions  $h_i(t), i = 1, \dots, m$  are log-concave implies that the functions  $1 - F_i(t)$  are log-concave as well. Moreover, using the log-concavity of  $1 - F_i(t)$ , we can show that  $g_i(t)$  is a decreasing function. Manipulating (2.44) we obtain

$$1 - F_i(y_i) = e^{-\int_{-\infty}^{y_i} g_i(t) dt}.$$

The functions  $g_i(t)$  are estimated from samples.

Let  $g_i^{(N)}$  denote an original estimator of  $g_i$  for a given  $N$ . We take a sample  $\{\mathbb{Z}_{Ni}\}$  from the population with distribution function  $F_i$ , and create a grid  $t_{N,1} < t_{N,2} < \dots < t_{N,m}$ . The original estimator  $g_i^{(N)}$  can then be defined as

$$g_i^{(N)}(t) = \frac{F_i^{(N)}(t_{N,j+1}) - F_i^{(N)}(t_{N,j})}{(t_{N,j+1} - t_{N,j})(1 - F_i^{(N)}(t_{N,j}))}, \quad t_{N,j} < t \leq t_{N,j+1},$$

where  $F_i^{(N)}$  is the empirical distribution function corresponding to  $F_i, i = 1, \dots, r$ .

We choose a point  $x_{N,j}$  from the interval  $(t_{N,j}, t_{N,j+1}]$  and a weight  $w(x_{N,j})$  associated with it. Then we solve the problem

$$\begin{aligned} & \inf_{U_j} \sum_{j=1}^m (U_j - g_i^{(N)}(x_{N,j}))^2 w(x_{N,j}) \\ & \text{subject to } U_j \geq U_{j+1}, \quad j = 1, \dots, m-1. \end{aligned}$$

Let  $\hat{g}_i^{(N)}(x_{N,j})$  be the optimal solution of this problem. We construct  $\hat{g}_i^{(N)}(\cdot)$  as a nondecreasing step function assigning the optimal solution to all arguments in the interval  $(t_{N,j}, t_{N,j+1}]$ . Further, we construct the approximation to  $F_i(y_i)$  by setting

$$\hat{F}_i^{(N)}(y_i) = 1 - e^{-\int_{-\infty}^{y_i} \hat{g}_i^{(N)}(t) dt}.$$

Now let us observe that the function

$$\log(1 - \hat{F}_i^{(N)}(y_i)) = -\int_{-\infty}^{y_i} \hat{g}_i^{(N)}(t) dt,$$

is piecewise linear and concave. Assume that the function consists of a finite number  $J_i$  of linear pieces given by the following equations:

$$a_{ij}^T z + b_{ij}, \quad j = 1, \dots, J_i, \quad i = 1, \dots, m.$$

Problem (2.43) is equivalent to the following linear programming problem:

$$\begin{aligned} & \min \langle c, x \rangle \\ & \text{subject to } z_i \geq a_{ij}^T z + b_{ij}, \quad j = 1, \dots, J_i \\ & \quad T_i x = z_i, \quad i = 1, \dots, m \\ & \quad Ax = b \\ & \quad x \geq 0. \end{aligned}$$

The solution of the latter problem is an approximate solution of the original problem.

### 2.5.6 A Response Surface Method

The method will be described for problem (2.26) under the assumption that the random vector  $Z$  has a continuous and log-concave distribution. This implies that the constraining function

$$G(x) = \mathbb{P}(g_1(x, Z) \geq 0, \dots, g_m(x, Z) \geq 0)$$

is log-concave in  $x$ . The idea of the method is to approximate  $G(\cdot)$  by a concave quadratic function  $Q(x) = x^T T x + h^T x + q$  ( $T$  is negative definite), then solve the approximate problem, take a new feasible point, improve the approximation, solve the problem with the new approximation *etc.* One difficulty is to develop a stopping rule in order to decide whether a solution is acceptable as an optimal solution. Some ideas are discussed in [102]. The algorithm can be described as follows.

Step 1. Given a collection of points  $J^k = \{x^0, \dots, x^{k-1}\}$  of feasible points and their corresponding values  $p_i = \log G(x^i)$ ,  $i = 1, \dots, k-1$ , solve the least squares problem

$$\min \sum_{i=0}^{k-1} (p_i - \langle x^i, T^k x^i \rangle + \langle h^k x \rangle + q^k)^2.$$

with respect to  $T^k, h^k, q^k$  such that  $T^k$  is negative semi-definite.

Step 2. We construct a quadratic function

$$\langle x^i, T^k x^i \rangle + \langle h^k x \rangle + q^k$$

and solve the approximate problem

$$\begin{aligned}
 & \min \langle c, x \rangle \\
 & \text{subject to } \langle x^i, T^k x^i \rangle + \langle h^k x \rangle + q^k \geq p \\
 & \quad Ax = b \\
 & \quad x \geq 0.
 \end{aligned}$$

Let  $x^k$  be an optimal solution.

Step 3. Check the stopping rule for  $x^k$ , and accept it as optimal solution, or return to Step 1.

### 2.5.7 Discrete Distribution

A straightforward way to solve problem (2.4) when  $Z$  has a discrete distribution is to find all  $p$ -efficient points and to process all corresponding problems (2.27) (see for example [282]). Specialized bounding-pruning techniques can be used to avoid solving all of them. For example, any feasible solution  $(\tilde{u}, \tilde{w})$  of the dual (2.28) can be used to generate a lower bound for (2.27). If it is worse than the best solution found so far, we can delete the problem (2.27); otherwise it has to be included into a list of problems to be solved exactly.

For multi-dimensional random vectors  $Z$  the number of  $p$ -efficient points can be very large and their straightforward enumeration – very difficult. It would be desirable, therefore, to avoid the complete enumeration and to search for promising  $p$ -efficient points only. This is accomplished by the next method.

### The cone generation method

This is a specialized method which uses the specificity of the discrete distributions. It is related to column generation methods, which have been known since the classical work [128] as extremely useful tools of large scale linear and integer programming [30, 89]. The method is based on the same idea as the primal-dual method for nonlinear constraints.

The algorithm works as follows:

Step 0. Select a  $p$ -efficient point  $v^0$ . Set  $J_0 = \{0\}$ ,  $k = 0$ .

Step 1. Solve the *master problem*

$$\min \langle c, x \rangle \tag{2.45}$$

$$Ax \geq b,$$

$$Tx \geq \sum_{j \in J_k} \lambda_j v^j, \tag{2.46}$$

$$\sum_{j \in J_k} \lambda_j = 1,$$

$$x \geq 0, \lambda \geq 0. \tag{2.47}$$

Let  $u^k$  be the vector of simplex multipliers associated with the constraint (2.46).

Step 2. Calculate an upper bound for the dual functional

$$\bar{d}(u^k) = \min_{j \in J_k} \langle u^k, v^j \rangle.$$

Step 3. Find a  $p$ -efficient solution  $v^{k+1}$  of the subproblem

$$\min_{z \in \mathcal{Z}_p} \langle u^k, z \rangle$$

and calculate

$$d(u^k) = \langle v^{k+1}, u^k \rangle.$$

Step 4. If  $d(u^k) = \bar{d}(u^k)$  then stop; otherwise set  $J_{k+1} = J_k \cup \{k+1\}$ , increase  $k$  by one and go to Step 1.

The first  $p$ -efficient point  $v^0$  can be found by solving the subproblem in Step 3, for an arbitrary  $u \geq 0$ . All master problems will be solvable, if the first one is solvable, *i.e.*, if the set  $\{x \in \mathbb{R}_+^n : Ax \geq b, Tx \geq v^0\}$  is non-empty. If not, adding a penalty term  $M\mathbb{1}^T t$  to the objective, and replacing (2.46) by

$$Tx + t \geq \sum_{j \in J_k} \lambda_j v^j,$$

with  $t \geq 0$  and a very large  $M$ , is the usual remedy ( $\mathbb{1}^T = [1 \ 1 \ \dots \ 1]$ ). The calculation of the upper bound at Step 2 is easy, because one can simply select  $j_k \in J_k$  with  $\lambda_{j_k} > 0$  and set  $\bar{d}(u^k) = (u^k)^T v^{j_k}$ . At Step 3 one may search for  $p$ -efficient solutions only, due to Lemma 5.

The algorithm is finite. Indeed, the set  $J_k$  cannot grow indefinitely, because there are finitely many  $p$ -efficient points (Theorem 17). If the stopping test of Step 4 is satisfied, optimality conditions for the convex hull problem (2.17) are satisfied. Moreover  $\hat{J}_k = \{j \in J_k : \langle v^j, u^k \rangle = d(u^k)\} \subseteq \hat{J}(u)$ .

When the dimension of  $x$  is large and the number of rows of  $T$  small, an attractive alternative to the cone generation method is provided by *bundle methods* applied directly to the dual problem

$$\max_{u \geq 0} [h(u) + d(u)],$$

because at any  $u \geq 0$  subgradients of  $h$  and  $d$  are readily available. For a comprehensive description of bundle methods the reader is referred to [161, 188].

Let us now focus our attention on solving the auxiliary problem in Step 3, which is explicitly written as

$$\min\{\langle u, z \rangle \mid F(z) \geq p\}, \quad (2.48)$$

where  $F(\cdot)$  denotes the distribution function of  $Z$ .

Assume that the components  $Z_i$ ,  $i = 1, \dots, s$ , are independent. Then we can write the probabilistic constraint in the following form:



$$\ln(F(z)) = \sum_{i=1}^s \ln(F_i(z_i)) \geq \ln p.$$

Since we know that at least one of the solutions is a  $p$ -efficient point, with no loss of generality we may restrict the search to grid vectors  $z$ . Furthermore, by Lemma 3, we have  $z_i \geq l_i$ , where  $l_i$  are  $p$ -efficient points of  $Z_i$ . For integer grids we obtain a nonlinear knapsack problem:

$$\begin{aligned} \min \quad & \sum_{i=1}^s u_i z_i \\ & \sum_{i=1}^s \ln(F_i(z_i)) \geq \ln p, \\ & z_i \geq l_i, \quad z_i \in \mathbb{Z}, \quad i = 1, \dots, s. \end{aligned}$$

If  $b_i$  is a known upper bound on  $z_i$ ,  $i = 1, \dots, s$ , we can transform the above problem to a 0–1 linear programming problem:

$$\begin{aligned} \min \quad & \sum_{i=1}^s \sum_{j=l_i}^{b_i} j u_i y_{ij} \\ & \sum_{i=1}^s \sum_{j=l_i}^{b_i} \ln(F_i(j)) y_{ij} \geq \ln p, \\ & \sum_{j=l_i}^{b_i} y_{ij} = 1, \quad i = 1, \dots, s, \\ & y_{ij} \in \{0, 1\}, \quad i = 1, \dots, s, \quad j = l_i, \dots, b_i. \end{aligned}$$

In this formulation,  $z_i = \sum_{j=l_i}^{b_i} j y_{ij}$ .

For log-concave marginals  $F_i(\cdot)$  the following compact formulation is possible. Setting  $z_i = l_i + \sum_{j=l_i+1}^{b_i} \delta_{ij}$  with binary  $\delta_{ij}$ , we can reformulate the problem as a 0–1 knapsack problem:

$$\begin{aligned} \min \quad & \sum_{i=1}^s \sum_{j=l_i+1}^{b_i} u_i \delta_{ij} \\ & \sum_{i=1}^s \sum_{j=l_i+1}^{b_i} a_{ij} \delta_{ij} \geq r, \\ & \delta_{ij} \in \{0, 1\}, \quad i = 1, \dots, s, \quad j = l_i + 1, \dots, b_i, \end{aligned}$$

where  $a_{ij} = \ln F_i(j) - \ln F_i(j-1)$  and  $r = \ln p - \ln F(l)$ . Indeed, by the log-concavity, we have  $a_{i,j+1} \leq a_{ij}$ , so there is always a solution with nonincreasing  $\delta_{ij}$ ,  $j = l_i + 1, \dots, b_i$ . Very efficient solution methods exist for such knapsack problems [239].

If the grid  $\mathcal{Z}$  is not integer we can map it to integers by numbering the possible realizations of each  $Z_i$  in an increasing order.

One advantage of the cone generation method is that we can separate the search for new  $p$ -efficient points (via (2.48)) and the solution of the ‘easy’ part of the problem: the master problem (2.45)–(2.47) in Step 1. Another advantage is that we do not need to generate and keep all  $p$ -efficient points.

Let us consider the optimal solution  $x^{\text{low}}$  of the convex hull problem (2.17) and the corresponding multipliers  $\lambda_j$ . Define  $J^{\text{low}} = \{j \in J : \lambda_j > 0\}$ .

If  $J^{\text{low}}$  contains only one element, the point  $x^{\text{low}}$  is feasible and therefore optimal for the disjunctive formulation (2.8). If, however, there are more positive  $\lambda$ ’s, we need to generate a feasible point. A natural possibility is to consider the *restricted disjunctive* formulation:

$$\begin{aligned} & \min \langle c, x \rangle \\ & \text{subject to } Tx \in \bigcup_{j \in J^{\text{low}}} K_j, \\ & \quad x \in \mathcal{D}. \end{aligned} \tag{2.49}$$

It can be solved by simple enumeration of all cases for  $j \in J^{\text{low}}$ :

$$\begin{aligned} & \min \langle c, x \rangle \\ & \text{subject to } Tx \geq v^j, \\ & \quad x \in \mathcal{D}. \end{aligned} \tag{2.50}$$

In general, it is not guaranteed that any of these problems has a non-empty feasible set. To ensure that problem (2.49) has a solution, it is sufficient that the following stronger version of Assumption 2.1 holds true.

**Assumption 2.2.** *The set  $\Lambda := \{(u, w) \in \mathbb{R}_+^{m+s} \mid A^T w + T^T u \leq c\}$  is non-empty and bounded.*

Indeed, then each of the dual problems (2.28) has an optimal solution, so by duality in linear programming each of the subproblems (2.50) has an optimal solution. We can, therefore, solve all of them and choose the best solution.

An alternative strategy would be to solve the corresponding upper bounding problem (2.50) every time a new  $p$ -efficient point is generated. If  $U_j$  denotes the optimal value of (2.50), the upper bound at iteration  $k$  is

$$\bar{U}^k = \min_{0 \leq j \leq k} U_j. \tag{2.51}$$

This may be computationally efficient, especially if we solve the dual problem (2.28), in which only the objective function changes from iteration to iteration.

If the distribution function of  $Z$  is  $\alpha$ -concave on the set of possible values of  $Z$ , Theorem 18 provides an alternative formulation of the upper bound problem (2.41):

$$\begin{aligned}
 & \min \quad \langle c, x \rangle \\
 & \text{subject to} \quad x \in \mathcal{D} \\
 & \quad Tx \geq z, \\
 & \quad z \in \mathbb{Z}^m, \\
 & \quad z \geq \sum_{j \in J_k} \lambda_j v^j, \\
 & \quad \sum_{j \in J_k} \lambda_j = 1 \\
 & \quad \lambda_j \geq 0, j \in J_k.
 \end{aligned} \tag{2.52}$$

Problem (2.52) provides a more accurate bound than the bound (2.51), because the set of integer  $z$  dominated by convex combinations of  $p$ -efficient points in  $J_k$  is not smaller than  $J_k$ . In fact, we need to solve this problem only at the end, with  $J_k$  replaced by  $J^{\text{low}}$ .

Special algorithms for probabilistic set-covering problem are presented in [39]. Branch-and-Bound techniques are developed in [40] for the case when  $x$  is an integer vector. The methods use the algebraic description of the feasible set by  $p$ -efficient points and suggest different techniques for generating the relevant  $p$ -efficient points.

### Bounds via binomial moments

If the components of  $Z$  are dependent it is difficult to evaluate the constraint function  $G(\cdot)$ , *e.g.*, for solving the subproblem (2.14) in the cone generation algorithm. Still, some bounds on its optimal solution may prove useful. A number of bounds are developed using only partial information on the distribution function of  $Z$  in the form of the marginal distributions:

$$F_{i_1 \dots i_k}(z_{i_1}, \dots, z_{i_k}) = \mathbb{P}\{Z_{i_1} \leq z_{i_1}, \dots, Z_{i_k} \leq z_{i_k}\}, \quad 1 \leq i_1 < \dots < i_k \leq m.$$

Since for each marginal distribution one has  $F_{i_1 \dots i_k}(z_{i_1}, \dots, z_{i_k}) \geq F(z)$  the following relaxation of  $\mathcal{Z}$  (defined by (2.5)) can be obtained.

**Lemma 8.** *For each  $z \in \mathcal{Z}_p$  and for every  $1 \leq i_1 < \dots < i_k \leq s$  the following inequality holds true:*

$$F_{i_1 \dots i_k}(z_{i_1}, \dots, z_{i_k}) \geq p.$$

We can determine probability bounds by solving certain linear optimization problem (see ([55, 279, 280])). The following result is known:

**Theorem 24.** *For any distribution function  $F : \mathbb{R}^m \rightarrow [0, 1]$  and any  $1 \leq k \leq m$ , at every  $z \in \mathbb{R}^m$  the optimal value of the following linear programming problem:*

$$\begin{aligned}
& \max \quad v_m \\
v_0 + v_1 + v_2 + \quad v_3 \quad + \cdots + \quad v_m &= 1 \\
v_1 + 2v_2 + \quad 3v_3 \quad + \cdots + \quad mv_m &= \sum_{1 \leq i \leq m} F_i(z_i) \\
v_2 + \quad \binom{3}{2}v_3 \quad + \cdots + \binom{m}{2}v_m &= \sum_{1 \leq i_1 < i_2 \leq m} F_{i_1 i_2}(z_{i_1}, z_{i_2}) \\
& \vdots \\
v_k + \binom{k+1}{k}v_{k+1} + \cdots + \binom{m}{k}v_m &= \sum_{1 \leq i_1 < \cdots < i_k \leq m} F_{i_1 \dots i_k}(z_{i_1}, \dots, z_{i_k}) \\
v_0 \geq 0, \quad v_1 \geq 0, \quad \dots, \quad v_m \geq 0.
\end{aligned} \tag{2.53}$$

provides an upper bound for  $F(z_1, \dots, z_m)$ .

We can use this result to bound the objective function in problem (2.48).

**Proposition 1.** *Let  $Z = (Z_1, \dots, Z_m)$  be an integer random vector and let  $F_{i_1, \dots, i_k}$  denote its marginal distribution functions. Then for every  $p \in (0, 1)$  and for every  $1 \leq k \leq m$  the optimal value of the problem*

$$\begin{aligned}
& \min \quad \langle u, z \rangle \\
v_0 + v_1 + v_2 + \quad v_3 \quad + \cdots + \quad v_m &= 1 \\
v_1 + 2v_2 + \quad 3v_3 \quad + \cdots + \quad mv_m &= \sum_{1 \leq i \leq m} F_i(z_i) \\
v_2 + \quad \binom{3}{2}v_3 \quad + \cdots + \binom{m}{2}v_m &= \sum_{1 \leq i_1 < i_2 \leq m} F_{i_1 i_2}(z_{i_1}, z_{i_2}) \\
& \vdots \\
v_k + \binom{k+1}{k}v_{k+1} + \cdots + \binom{m}{k}v_m &= \sum_{1 \leq i_1 < \cdots < i_k \leq m} F_{i_1 \dots i_k}(z_{i_1}, \dots, z_{i_k}) \\
v_0 \geq 0, \quad v_1 \geq 0, \quad \dots, \quad v_{s-1} \geq 0, \quad v_m \geq p, \quad z_1 \geq l_1, \quad z_2 \geq l_2, \quad \dots, \quad z_m \geq l_m, \\
& z \in \mathbb{Z}^m
\end{aligned} \tag{2.54}$$

provides a lower bound on the optimal value  $d(u)$  given by (2.48).

**Proof.** If  $z \in \mathcal{Z}$ , that is,  $F(z) \geq p$ , then the optimal value of (2.53) satisfies  $v_m \geq p$ . Thus  $z$  and the solution  $v$  of (2.53) are feasible for (2.54). Since the objective functions of (2.48) and (2.54) are the same, the result follows.  $\square$

Problem (2.54) is a nonlinear mixed-integer problem. Its advantage over the original formulation is that it uses marginal functions in an explicit way which allows for the development of specialized solution methods.

### 2.5.8 Probabilistic Valid Inequalities

A relation between probabilistic constraints and the theory of valid inequalities in integer programming has been developed in [311].

We shall just sketch some ideas in this direction. Let us assume that the distribution of  $Z$  is approximated by finitely many scenarios  $z_1, \dots, z_S$  having probabilities  $p_1, \dots, p_S$ . Under mild assumptions problem (2.1) can be converted to a mixed-integer programming problem

$$\min f(x) \tag{2.55}$$

$$\text{subject to } g(x, z_s) + Mv_s \geq 0, \quad s = 1, \dots, S, \tag{2.56}$$

$$\sum_{s=1}^S p_s v_s \leq 1 - p, \tag{2.57}$$

$$\begin{aligned} x &\in \mathcal{D}, \\ v_s &\in \{0, 1\}, \quad s = 1, \dots, S, \end{aligned} \tag{2.58}$$

where  $M$  is a vector with sufficiently large components, so  $v_s = 1$  makes (2.56) trivial. Each binary variable  $v_s$  indicates whether the current solution  $x$  violates the constraint  $g(x, z_s) \geq 0$  or not, and the probability constraint takes on the form of the knapsack inequality (2.57).

In many applications it is possible to determine a partial order ' $\preceq$ ' in the set of scenarios  $z_s$ , representing their difficulty for the constraints  $g(x, z_s) \geq 0$ . In the simplest setting, we shall have  $z_s \preceq z_\sigma$  ( $z_s$  is easier than  $z_\sigma$ ) if

$$g(x, z_\sigma) \geq 0 \Rightarrow g(x, z_s) \geq 0, \quad \text{for all } x \in X.$$

Then the mixed-integer formulation (2.55)–(2.58) can be augmented with the *precedence constraint*:

$$v_s \leq v_\sigma \text{ if } z_s \preceq z_\sigma.$$

Probabilistic valid inequalities for the binary variables  $v_s$  are developed on the basis of this structure. For each scenario  $z_s$  we define the set of comparable scenarios which are at least as hard as  $z_s$ :

$$A_s = \{z_j : z_s \preceq z_j\}.$$

If we fail to satisfy the constraint  $g(x, z_s) \geq 0$  for scenario  $s$ , we shall fail for all scenarios in  $A_s$ , *i.e.*,  $v_j = 1$  for all  $z_j \in A_s$ . In this way probabilistic counterparts of the concepts of a cover and cover inequalities known from integer programming are introduced. In our setting, a set  $C \subset \{1, \dots, S\}$  is an *induced cover* if

$$\mathbb{P} \left\{ \bigcup_{s \in C} A_s \right\} > 1 - p.$$

If  $v_s = 1$  for all  $s \in C$ , then we must have  $v_j = 1$  for all  $z_j$  in the union of the sets  $A_s$ ,  $s \in C$ , and the probability constraint (2.57) will be violated. Therefore the following *induced cover inequality* must hold true:

$$\sum_{s \in C} v_s \leq |C| - 1.$$

The second implication of the partial order is that we do not need to enforce inequality (2.56) for all scenarios. By a similar argument, if  $g(x, z_s) \geq 0$ , then  $g(x, z_\sigma) \geq 0$  for all  $z_\sigma \preceq z_s$ . Thus, we may try to determine a set  $\mathcal{L}$  of *critical scenarios*, similarly to the set of  $p$ -efficient points of the problem with the random right hand side.

These two basic ideas can be put together to formulate the following approximate problem:

$$\begin{aligned}
& \min f(x) \\
& \text{subject to } g(x, z_s) + Mv_s \geq 0 \quad s \in \mathcal{L}, \\
& \sum_{s=1}^S p_s v_s \leq 1 - p, \\
& x \in X, \\
& v_s \leq v_\sigma, \quad \text{if } z_s \preceq z_\sigma \\
& v_s \in [0, 1], \quad s = 1, \dots, S, \\
& \sum_{s \in C} v_s \leq |C| - 1.
\end{aligned}$$

For a detailed description of this solution technique we refer to [311].

## 2.6 Cash Matching with Probabilistic Liquidity Constraints

There are many publications addressing interesting applications of probabilistic constraints. We do not attempt to address the potential of probabilistic optimization for solving applied problems. We return to a version of our starting example because our duality theory finds an interesting interpretation in its context.

We consider the following cash matching problem. We have random liabilities  $L_t$  in periods  $t = 1, \dots, T$  and a basket of  $n$  bonds. The payment of bond  $i$  in period  $t$  is denoted by  $a_{it}$ . It is zero for  $t$  before the purchase of the bond and for  $t$  greater than the maturity time of the bond. At the time of purchase  $a_{it}$  is the negative of the price of the bond, at the following periods it is equal to the coupon payment, and at the time of maturity it is equal to the face value plus the coupon payment. Our initial capital equals  $c_0$ .

The objective is to design a bond portfolio such that the probability of covering the liabilities over the entire period  $1, \dots, T$  is at least  $p$ . Subject to this condition, we want to maximize the final cash on hand, guaranteed with probability  $p$ .

Let us introduce the cumulative liabilities

$$Z_t = \sum_{\tau=1}^t L_\tau, \quad t = 1, \dots, T.$$

Denoting by  $x_i$  the amount invested in bond  $i$ , we observe that the cumulative cash flows up to time  $t$ , denoted  $c_t$ , can be expressed as follows:

$$c_t = c_{t-1} + \sum_{i=1}^n a_{it}x_i, \quad t = 1, \dots, T.$$

Using cumulative cash flows and cumulative liabilities permits the carry-over of capital from one stage to the next one, while keeping the random quantities at the right hand side of the constraints. The problem takes on the form

$$\begin{aligned} & \max c_T \\ & \text{subject to } \mathbb{P}[c_t \geq Z_t, \quad t = 1, \dots, T] \geq p, \\ & \quad c_t = c_{t-1} + \sum_{i=1}^n a_{it}x_i, \quad t = 1, \dots, T, \\ & \quad x \geq 0. \end{aligned}$$

Let us observe that first constraint of this problem is a probabilistic liquidity constraint. If the vector  $Z$  has a quasi-concave distribution (*e.g.*, joint normal distribution), the resulting problem is convex. Thus both the dual method from Section 2.5.3 and the primal-dual method from Section 2.5.4 yield optimal solutions of the problem.

The convex hull problem (2.17) can be now written as follows:

$$\begin{aligned} & \max c_T \\ & \text{subject to } c_t = c_{t-1} + \sum_{i=1}^n a_{it}x_i, \quad t = 1, \dots, T, \end{aligned} \quad (2.59)$$

$$c_t \geq \sum_{j=1}^{T+1} \lambda_j v_t^j, \quad t = 1, \dots, T, \quad (2.60)$$

$$\sum_{j=1}^{T+1} \lambda_j = 1, \quad (2.61)$$

$$\lambda \geq 0, \quad x \geq 0. \quad (2.62)$$

In constraint (2.60) the vectors  $v^j = (v_1^j, \dots, v_T^j)$ , for  $j = 1, \dots, T+1$ , are  $p$ -efficient trajectories of the cumulative liabilities  $Z = (Z_1, \dots, Z_T)$ . Constraints (2.60)–(2.62) require that the cumulative cash flows are greater than or equal to a convex combination of  $p$ -efficient trajectories. Recall that by Lemma 4, no more than  $T+1$   $p$ -efficient trajectories are needed. Unfortunately, we do not know the optimal collection of these trajectories.

Let us assign non-negative Lagrange multipliers  $u = (u_1, \dots, u_T)$  to the constraint (2.60), multipliers  $w = (w_1, \dots, w_T)$  to the constraints (2.59), and a multiplier  $\rho \in \mathbb{R}$  to the constraint (2.61). For the convenience of notation we introduce the constant  $w_{T+1} = 1$ . The dual problem becomes

$$\min c_0 w_1 - \rho \quad (2.63)$$

$$\text{subject to } w_t = w_{t+1} + u_t, \quad t = T, T-1, \dots, 1, \quad (2.64)$$

$$\sum_{t=1}^T w_t a_{it} \leq 0, \quad i = 1, \dots, n, \quad (2.65)$$

$$\rho \leq \sum_{t=1}^T u_t v_t^j, \quad j = 1, \dots, T+1. \quad (2.66)$$

We can observe that each dual variable  $u_t$  is the cost of borrowing a unit of cash for one time period,  $t$ . The amount  $u_t$  is to be paid at the end of the planning horizon. It follows from (2.64) that each multiplier  $w_t$  is the amount that has to be returned at the end of the planning horizon if a unit of cash is borrowed at  $t$  and held till  $T$ .

The constraints (2.65) represent the *non-arbitrage condition*. For each bond  $i$  we can consider the following operation: borrow money to buy the bond and lend away its coupon payments, according to the rates implied by  $w_t$ 's. At the end of the planning horizon, we collect all loans and pay off the debt. The profit from this operation should be non-positive for each bond, and this is represented by (2.65).

Let us observe that each product  $u_t v_t^j$  is the the amount that has to be paid at the end, for having a debt in the amount  $v_t^j$  in period  $t$ . Recall that  $v_t^j$  is the  $p$ -efficient cumulative liability up to time  $t$ . Denote the implied one-period liabilities by

$$\begin{aligned} L_t^j &= v_t^j - v_{t-1}^j, \quad t = 2, \dots, T, \\ L_1^j &= v_1^j. \end{aligned}$$

Changing the order of summation, we obtain

$$\sum_{t=1}^T u_t v_t^j = \sum_{t=1}^T u_t \sum_{\tau=1}^t L_\tau^j = \sum_{\tau=1}^T L_\tau^j \sum_{t=\tau}^T u_t = \sum_{\tau=1}^T L_\tau^j (w_\tau - 1).$$

It follows that the sum appearing at the right hand side of (2.66) is the extra cost of covering the  $j$ th  $p$ -efficient liability sequence by borrowed money, that is, the difference between the amount that has to returned at the end of the planning horizon, and the total liability. The variable  $\rho$ , therefore, represents the minimal cost of this form, for all  $p$ -efficient trajectories. This allows us to interpret the dual objective function (2.63) as the amount obtained at  $T$  for lending away our capital  $c_0$  decreased by the extra cost of covering a  $p$ -efficient liability sequence by borrowed money. By duality this quantity is the same as  $c_T$ , which implies that both ways of covering the liabilities are equally profitable.

To observe the work of the methods we have used data on 72 government bonds and AAA corporate bonds ranging from 6-month treasury bills (which



do not pay coupons, but sell at discount) to 5-year bonds paying coupons each 6-months. The liabilities were assumed to be normally distributed with expectation 2,000,000 and standard deviation 100,000. The initial capital was  $c_0 = 20,000,000$  and the number of 6-month periods  $T = 10$ . The probability  $p = 0.95$ . To facilitate the numerical solution of the method, the distribution of the liabilities was approximated by  $N = 100$  equally likely scenarios.

The dual and the primal-dual methods were used to solve the problem. The search for new  $p$ -efficient points in both methods was implemented as a simple binary optimization problem with a knapsack constraint. Other subproblems were solved by the CPLEX linear programming solver.

The dual method terminated after 34 iterations finding the optimal portfolio of 9 bonds of different maturities. The primal-dual method found exactly the same solution after just 3 iterations. In both cases the computation time on a 1.7GHz PC was less than one minute.

The key element of both methods is the subproblem for generating  $p$ -efficient points.

The problem at hand was linear, and therefore both methods were equally easy to implement. If the functions  $f$  and  $g$  are nonlinear, one iteration of the primal-dual method requires more computational effort than one iteration of the dual method.

*Acknowledgement.* This research is supported by the NSF awards 0354500 and 0303728.

Probabilistic and Randomized Methods for Design under  
Uncertainty

Calafiore, G.; Dabbene, F. (Eds.)

2006, XIV, 458 p., Hardcover

ISBN: 978-1-84628-094-8