Kybernetika

Daniel Kuhn; Panos Parpas; Berç Rustem
Bound-based decision rules in multistage stochastic programming

Kybernetika, Vol. 44 (2008), No. 2, 134--150

Persistent URL: http://dml.cz/dmlcz/135840

Terms of use:

© Institute of Information Theory and Automation AS CR, 2008

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.

This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* http://project.dml.cz

KYBERNETIKA — VOLUME 44 (2008), NUMBER 2, PAGES 134 – 150

BOUND–BASED DECISION RULES IN MULTISTAGE STOCHASTIC PROGRAMMING

Daniel Kuhn, Panos Parpas and Berç Rustem

We study bounding approximations for a multistage stochastic program with expected value constraints. Two simpler approximate stochastic programs, which provide upper and lower bounds on the original problem, are obtained by replacing the original stochastic data process by finitely supported approximate processes. We model the original and approximate processes as dependent random vectors on a joint probability space. This probabilistic coupling allows us to transform the optimal solution of the upper bounding problem to a near-optimal decision rule for the original problem. Unlike the scenario tree based solutions of the bounding problems, the resulting decision rule is implementable in all decision stages, i. e., there is no need for dynamic reoptimization during the planning period. Our approach is illustrated with a mean-risk portfolio optimization model.

Keywords: stochastic programming, bounds, decision rules, expected value constraints, portfolio optimization

AMS Subject Classification: 90C15

1. INTRODUCTION

Consider a sequential decision problem under uncertainty. Decisions are selected at different stages indexed by $h = 1, \ldots, H$, and the uncertainty is given in terms of two stochastic processes $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_H)$ and $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_H)$. All random objects (that is, random variables, random vectors, and stochastic processes) are defined on some sample space (Ω, Σ, P). They are consistently denoted in boldface, while their realizations are denoted by the same symbols in normal script. We assume that the random vectors $\boldsymbol{\eta}_h$ and $\boldsymbol{\xi}_h$ take values in $\mathbb{R}^K$ and $\mathbb{R}^L$, respectively. The process $\boldsymbol{\eta}$ will impact only the objective function, and $\boldsymbol{\xi}$ will influence only the constraints of our decision problem. This convention is nonrestrictive since a random process that appears both in the objective and the constraints can be duplicated in both $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$.[1] At stage h the decision maker has full information about the history of the stochastic processes up to stage h but no (or only distributional) information about the future from stage $h+1$ onwards. Hence, decisions $\boldsymbol{x}_h$ selected at stage h represent

[1]However, common components of $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ will be treated differently in Section 3 when we address their discretization.

$\mathcal{F}_h$-measurable random vectors valued in $\mathbb{R}^N$, where $\mathcal{F}_h$ is the σ-algebra generated by $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_h$ and $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_h$. Since all random objects in this paper are defined on (Ω, Σ, P), the induced σ-algebras $\mathcal{F}_h$ are subsets of Σ. A similar statement holds for all other σ-algebras to be introduced below. By construction, the decision process $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_H)$ is adapted to $\mathbb{F} = \{\mathcal{F}_h\}_{h=1}^H$, that is, the filtration induced by the data processes $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$. This property is usually referred to as *non-anticipativity* in literature.

In this paper we study a class of multistage stochastic programs with expected value constraints.

$$\begin{aligned} &\underset{\boldsymbol{x} \in X(\mathbb{F})}{\text{minimize}}\ \mathrm{E}\left(c(\boldsymbol{x}, \boldsymbol{\eta})\right) \\ &\text{s.t.}\quad \mathrm{E}(f_h(\boldsymbol{x}, \boldsymbol{\xi}) \,|\, \mathcal{F}_h) \leq 0 \quad P\text{-a. s. } \forall\, h = 1, \ldots, H \end{aligned} \tag{$\mathcal{P}$}$$

The objective criterion and the constraints in $\mathcal{P}$ are determined through a real-valued cost function $c : \mathbb{R}^{H(N+K)} \to \mathbb{R}$ and a sequence of vector-valued constraint functions $f_h : \mathbb{R}^{H(N+L)} \to \mathbb{R}^M$ for $h = 1, \ldots, H$. We can assume without loss of generality that there are equally many decision variables and constraints in all stages. Minimization in $\mathcal{P}$ is over a space of stochastic decision processes, which are also referred to as strategies or policies. The set of admissible strategies is defined as

$$X(\mathbb{F}) = \{\boldsymbol{x} \in \times_{h=1}^H \mathcal{L}^\infty(\Omega, \mathcal{F}_h, P; \mathbb{R}^N) \,|\, \boldsymbol{x}(\omega) \in X \text{ for } P\text{-a.e. } \omega \in \Omega\},$$

where X is a convex compact subset of $\mathbb{R}^{HN}$. In problem $\mathcal{P}$ the constraint functions are required to be nonpositive in expectation (instead of pointwise), where expectation is conditional on the stagewise information sets. If some component functions of f_h depend only on $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_h$ and $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_h$, then the corresponding conditional expectations become vacuous, and we essentially face standard constraints that hold pointwise. Expected value constraints are important for modelling the risk attitude of the decision maker, e. g. through conditional value-at-risk constraints [28] or integrated chance constraints [12]. They also have relevance in Markowitz-type portfolio optimization problems which restrict the expectation of terminal wealth.

In the sequel, the following regularity conditions are assumed to hold:

(C1) c is convex in x, concave in η, and continuous;

(C2) there is a convex continuous function

$$\tilde{f}_h : \mathbb{R}^{(HN+1)\times(HL+1)} \to \mathbb{R}^M \quad \text{with} \quad f_h(x, \xi) = \tilde{f}_h\left((1, x)(1, \xi)^\top\right),$$

and $\tilde{f}_h$ is constant in $x_i\, \xi_j^\top$ for all $1 \leq j \leq i \leq H$, $h = 1, \ldots, H$;

(C3) X is convex and compact;

(C4) $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are serially independent processes, and there are compact convex sets $\Theta \subset \mathbb{R}^{HK}$ and $\Xi \subset \mathbb{R}^{HL}$ with $\boldsymbol{\eta}(\omega) \in \Theta$ and $\boldsymbol{\xi}(\omega) \in \Xi$ for P-a.e. $\omega \in \Omega$.

The conditions (C1) and (C2) are satisfied, for instance, by linear multistage stochastic programs which have deterministic recourse matrices but random objective function coefficients, right hand side vectors, and technology matrices. The compactness

requirement (C3) is nonrestrictive in practice since we want to solve $\mathcal{P}$ numerically. Reasonable numerical solutions are always bounded. A similar comment applies to condition (C4). The supports of $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ can always be compactified by truncating certain extreme scenarios that have a marginal effect on the solution of $\mathcal{P}$. The serial independence requirement is for technical reasons. It can often be circumvented by rewriting the original data process as a transformation of a serially independent noise process. In this case, any serial dependencies are absorbed in the definition of the cost and constraint functions.[2] Notice that we do *not* require problem $\mathcal{P}$ to have (relatively) complete recourse.

Following the lines of [21], we next introduce the set

$$Y(\mathbb{F}) = \{\boldsymbol{y} \in \times_{h=1}^{H} \mathcal{L}^1(\Omega, \mathcal{F}_h, P; \mathbb{R}^M) \,|\, \boldsymbol{y}(\omega) \geq 0 \text{ for } P\text{-a.e. } \omega \in \Omega\},$$

which is interpreted as the set of all nonnegative integrable *dual* decision processes $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_H)$ adapted to $\mathbb{F}$. By using $Y(\mathbb{F})$ to dualize the explicit constraints in $\mathcal{P}$, we obtain an equivalent min-max problem *without* explicit constraints. This statement is formalized in the following lemma.

Lemma 1. (Wright [31, § 4]) Under the conditions (C1) – (C3) we find

$$\inf \mathcal{P} = \inf_{\boldsymbol{x} \in X(\mathbb{F})} \sup_{\boldsymbol{y} \in Y(\mathbb{F})} \mathrm{E}\left(c(\boldsymbol{x}, \boldsymbol{\eta}) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot f_h(\boldsymbol{x}, \boldsymbol{\xi}) \right). \tag{1}$$

Unless the support of $(\boldsymbol{\eta}, \boldsymbol{\xi})$ constitutes a finite subset of $\mathbb{R}^{H(K+L)}$, problem $\mathcal{P}$ constitutes an infinite-dimensional optimization problem, which is computationally untractable. A simple way to address this problem is to replace the original data processes by finitely supported approximate processes $\boldsymbol{\eta}^u = (\boldsymbol{\eta}_1^u, \ldots, \boldsymbol{\eta}_H^u)$ and $\boldsymbol{\xi}^u = (\boldsymbol{\xi}_1^u, \ldots, \boldsymbol{\xi}_H^u)$ of appropriate dimensions. The resulting approximate problem $\mathcal{P}^u$ can be represented as

$$\begin{array}{l} \underset{\boldsymbol{x} \in X(\mathbb{F}^u)}{\text{minimize}} \; \mathrm{E}\left(c(\boldsymbol{x}, \boldsymbol{\eta}^u)\right) \\ \text{s.t.} \quad \mathrm{E}(f_h(\boldsymbol{x}, \boldsymbol{\xi}^u) \,|\, \mathcal{F}_h^u) \leq 0 \quad P\text{-a.s. } \forall\, h = 1, \ldots, H, \end{array} \tag{$\mathcal{P}^u$}$$

where $\mathbb{F}^u = \{\mathcal{F}_h^u\}_{h=1}^H$ stands for the filtration generated by $\boldsymbol{\eta}^u$ and $\boldsymbol{\xi}^u$, while $X(\mathbb{F}^u)$ is defined as the space of $\mathbb{F}^u$-adapted primal decision processes valued in X. As before, it proves useful to introduce a space $Y(\mathbb{F}^u)$ which contains all $\mathbb{F}^u$-adapted dual decision processes valued in the nonnegative orthant of $\mathbb{R}^M$. Notice that Lemma 1 also applies to problem $\mathcal{P}^u$. This result will help us to clarify the relation between the original and approximate stochastic optimization problems in Section 4 below.

[2] Besides the compact support and serial independence assumptions, we impose no further conditions on the distribution of the data process. Although we require the cost and constraint functions to be continuous in their arguments, the distribution of $(\boldsymbol{\eta}, \boldsymbol{\xi})$ can be discrete.

2. DECISION RULES

In stochastic programming the original and approximate data processes are mostly identified, while only their distributions are distinguished. In this mainstream view, the stochastic program $\mathcal{P}$ is approximated by discretizing the underlying probability measure, which is equivalent to generating a discrete *scenario tree*. A multitude of different methods has been suggested to perform this kind of scenario generation. Discrete approximate distributions can be constructed, for instance, by sampling from the original distribution [16, 17, 18], by using concepts of numerical integration [19, 25], by solving mass transportation problems [13, 14, 26, 27], or by solving certain moment problems [3, 4, 5, 6, 7, 10, 15].

The method advocated in this paper adopts a radically different perspective in that the original and approximate data processes are interpreted as two dependent random vectors on a single probability space and not as a single random vector under two alternative probability measures. Instead of specifying the *marginal* distribution of $(\boldsymbol{\eta}^u, \boldsymbol{\xi}^u)$ on $\mathbb{R}^{H(K+L)}$, we prescribe its *conditional* distributions given the original process $(\boldsymbol{\eta}, \boldsymbol{\xi})$. By using the product measure theorem [1, Theorem 2.6.2], this will allow us to determine the *joint* distribution of the original and approximate processes on $\mathbb{R}^{2H(K+L)}$.

At first sight, it seems cumbersome to deal with the rather complex joint distribution instead of simply focussing on the marginal distributions of the original and approximate data processes. This objection appears particularly compelling since only the marginal distributions of the approximate (original) data processes influence the solution of the approximate (original) stochastic program. The joint distribution is an artefact that has no 'physical' meaning. Nevertheless, we will argue below that it can be very beneficial to have this joint distribution at our disposition. It provides a convenient way to transform optimal solutions of the approximate problem to near-optimal solutions for the original problem.

It is frequently criticized that only the first-stage optimal solution of a scenario tree-based stochastic program is implementable. The reason for this is that recourse decisions are only available for a finite number of scenarios. Any scenario that materializes in reality will typically differ from all scenarios in the tree, and therefore the scenario tree solution provides no guidelines on what to do in reality at stages $h = 2, \ldots, H$. This difficulty can be overcome by using a *rolling optimization* scheme. That is, the approximate stochastic program is solved repeatedly at all stages h over a shrinking (or receding) planning horizon and conditional on the realizations of the random data observed up to h. In each step of this procedure, only the corresponding first-stage decision is implemented. Rolling optimization schemes of this type have been evaluated e. g. by Kouwenberg [20] and Fleten et al. [8]. Notice that the rolling optimization strategy is always implementable, while the optimal strategy of the approximate problem (for $h = 1$) is generically not implementable. It should also be emphasized that the solution of the approximate problem provides only limited information about the expected cost of the rolling optimization strategy. The latter could principally be estimated by sampling many paths of the data process $(\boldsymbol{\eta}, \boldsymbol{\xi})$, evaluating the rolling optimization strategy along each path, and averaging the pathwise cost over all samples. This involves solving H multistage stochastic programs

for each sample, which quickly becomes computationally prohibitive. Therefore, forecasting the expected cost of the rolling optimization strategy is at best difficult. Observe that rolling optimization is practicable, however, if one focusses only on one path of the data process $(\boldsymbol{\eta}, \boldsymbol{\xi})$, e. g. the path that materializes in reality.

In this paper we report on a simple method to transform the optimal solution of the approximate problem (for $h = 1$) to a near-optimal solution for the original problem. Like any admissible policy, this near-optimal solution can be represented in feedback form, that is, as a measurable function of (η, ξ). In standard terminology, such functions are referred to as *decision rules*. The decision rule obtained by our methodology will turn out to be piecewise constant in η and piecewise polynomial in ξ. We will further provide deterministic error bounds that estimate the optimality gap associated with our decision rule. Like the rolling optimization strategy mentioned above, our near-optimal decision rule is implementable in reality, that is, it prescribes an admissible sequence of actions for each scenario of the original data process. In contrast to the rolling optimization strategy, however, our decision rule is determined by the solution of one single stochastic program. Therefore, its expected cost (as well as the underlying cost distribution) can conveniently and accurately be evaluated via Monte Carlo simulation. No reoptimization is necessary at stages $h = 2, \ldots, H$ to achieve this. Our approach has the additional benefit that it accommodates expected value constraints, which frequently arise when dealing with risk functionals in the objective and or in the constraints. Moreover, our method is even applicable to problems that fail to have relatively complete recourse.

A review and critical assessment of early results on decision rules in stochastic programming is due to Garstka and Wets [11]. The use of parametric *linear* decision rules has recently been proposed by Shapiro and Nemirovski [29] to reduce the computational complexity of multistage stochastic programs. For similar reasons, Thénié and Vial [30] suggest piecewise constant *step decision rules*, which are obtained by using machine learning techniques. Mirkov and Pflug [24] construct near-optimal decision rules for multistage stochastic programs that have complete recourse and satisfy certain Lipschitz conditions.

3. DISCRETIZATION

As indicated in the previous section, $\boldsymbol{\eta}^u$ and $\boldsymbol{\xi}^u$ are constructed as finitely supported stochastic processes on (Ω, Σ, P), which are (nonlinearly) correlated with the original processes $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$. For the following discussion we specify the underlying probability space. Without loss of generality, we set $\Omega = \mathbb{R}^{2H(K+L)}$, let Σ be the Borel σ-algebra on Ω, and define $\boldsymbol{\eta}$, $\boldsymbol{\xi}$, $\boldsymbol{\eta}^u$, and $\boldsymbol{\xi}^u$ as the coordinate projections

$$
\begin{aligned}
\boldsymbol{\eta}: &\quad \mathbb{R}^{2H(K+L)} \to \mathbb{R}^{HK}, \quad (\eta, \xi, \eta^u, \xi^u) \mapsto \eta\,, \\
\boldsymbol{\xi}: &\quad \mathbb{R}^{2H(K+L)} \to \mathbb{R}^{HL}, \quad (\eta, \xi, \eta^u, \xi^u) \mapsto \xi\,, \\
\boldsymbol{\eta}^u: &\quad \mathbb{R}^{2H(K+L)} \to \mathbb{R}^{HK}, \quad (\eta, \xi, \eta^u, \xi^u) \mapsto \eta^u\,, \\
\boldsymbol{\xi}^u: &\quad \mathbb{R}^{2H(K+L)} \to \mathbb{R}^{HL}, \quad (\eta, \xi, \eta^u, \xi^u) \mapsto \xi^u\,.
\end{aligned}
$$

Furthermore, we let P be the joint distribution of these stochastic processes. Note that only the marginal distribution of $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ is a priori given through the specification of problem $\mathcal{P}$. In contrast, the conditional distribution of $(\boldsymbol{\eta}^u, \boldsymbol{\xi}^u)$ given $(\boldsymbol{\eta}, \boldsymbol{\xi})$ is selected at our discretion. Once we have specified this conditional distribution, the joint distribution P of the original and approximate stochastic processes follows from the product measure theorem [1, Theorem 2.6.2]. Strictly speaking, the probability measure P is therefore partially unknown before we have specified the conditional distribution of the approximate data processes. However, for the purposes of this article it is perfectly acceptable to assume that P is known (or chosen appropriately) already at the outset.

Our aim is to specify the distribution of $(\boldsymbol{\eta}^u, \boldsymbol{\xi}^u)$ conditional on $(\boldsymbol{\eta}, \boldsymbol{\xi})$ in such a way that the following relations are satisfied for suitable versions of the conditional expectations, respectively.

$$\mathrm{E}(\boldsymbol{x}|\mathcal{F}) \in X(\mathbb{F}) \text{ for all } \boldsymbol{x} \in X(\mathbb{F}^u) \tag{2a}$$

$$\mathrm{E}(\boldsymbol{y}|\mathcal{F}^u) \in Y(\mathbb{F}^u) \text{ for all } \boldsymbol{y} \in Y(\mathbb{F}) \tag{2b}$$

$$\mathrm{E}(\boldsymbol{\xi}^u|\mathcal{F}) = \boldsymbol{\xi} \tag{2c}$$

$$\mathrm{E}(\boldsymbol{\eta}|\mathcal{F}^u) = \boldsymbol{\eta}^u \tag{2d}$$

Here, we use the shorthand notation $\mathcal{F} = \mathcal{F}^H$ and $\mathcal{F}^u = \mathcal{F}^u_H$, both of which are sub-$\sigma$-algebras of Σ. Observe that the random parameters $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ and their discretizations are treated in a complementary manner in (2c) and (2d). The reason for this is that the problem Lagrangian $L(x, y; \eta, \xi) = c(x, \eta) + \sum_{h=1}^{H} y_h^\top f_h(x, \xi)$ is convex in ξ while it is concave in η; see also the proof of Theorem 4 below. Furthermore, we require the following inequality to hold for $h = 1, \ldots, H$.

$$\mathrm{E}(f_h(\boldsymbol{x}, \boldsymbol{\xi}^u)|\mathcal{F}) \geq f_h(\mathrm{E}(\boldsymbol{x}|\mathcal{F}), \boldsymbol{\xi}) \quad P\text{-a. s.} \quad \text{for all} \quad \boldsymbol{x} \in X(\mathbb{F}^u) \tag{3}$$

The relation (3) can be seen as a generalization of the conditional Jensen inequality which applies to constraint functions subject to the assumption (C2). A general approach to construct conditional distribution functions for $\boldsymbol{\eta}^u$ and $\boldsymbol{\xi}^u$ which guarantee the validity of (2) and (3) has been presented in [21, § 4]; see also [23, § 5]. Here, we will not repeat this construction in its full generality, but we will sketch the underlying ideas.

To keep things simple, we consider first the one-stage case and assume that there is only one random parameter in the constraints. Put differently, we assume that $H = 1$, $K = 0$, and $L = 1$. By the regularity condition (C4), the real-valued random variable $\boldsymbol{\xi} = \boldsymbol{\xi}_1$ is supported in a compact convex set $\Xi = [a, b] \subset \mathbb{R}$. The marginal distribution of $\boldsymbol{\xi}$ is assumed to be given, see Figure 1. In contrast, the conditional distribution of $\boldsymbol{\xi}^u$ given $\boldsymbol{\xi}$ is chosen with the aim to satisfy (2) and (3). Using an educated guess, we require this conditional distribution to be supported on the two extreme points of Ξ, and the two conditional probabilities are set to

$$P(\boldsymbol{\xi}^u = a \,|\, \boldsymbol{\xi} = \xi) = \frac{b - \xi}{b - a}, \quad \text{and} \quad P(\boldsymbol{\xi}^u = b \,|\, \boldsymbol{\xi} = \xi) = \frac{\xi - a}{b - a},$$

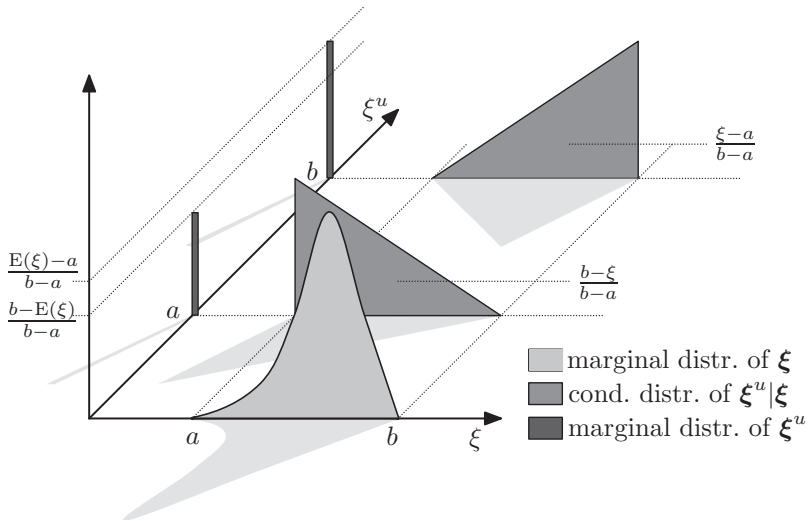

Fig. 1. Relation between $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^u$.

see Figure 1. By construction, the marginal distribution of $\boldsymbol{\xi}^u$ is concentrated to the points a and b with unconditional probabilities

$$P(\boldsymbol{\xi}^u = a) = \frac{b - \mathrm{E}(\boldsymbol{\xi})}{b-a}, \quad \text{and} \quad P(\boldsymbol{\xi}^u = b) = \frac{\mathrm{E}(\boldsymbol{\xi}) - a}{b-a}.$$

These probabilities play an important role in the well-known one-dimensional Edmundson–Madansky inequality, see e. g. [2, p. 293], which follows as a special case from Theorem 2 below. It is easy to verify that condition (2c) is satisfied since

$$\mathrm{E}(\boldsymbol{\xi}^u|\mathcal{F}) = a\,P(\boldsymbol{\xi}^u = a\,|\,\boldsymbol{\xi}) + b\,P(\boldsymbol{\xi}^u = b\,|\,\boldsymbol{\xi}) = a\,\frac{b-\boldsymbol{\xi}}{b-a} + b\,\frac{\boldsymbol{\xi}-a}{b-a} = \boldsymbol{\xi} \quad P\text{-a. s.}$$

The conditions (2a) and (2b) are trivially satisfied for $H = 1$, and condition (2d) becomes vacuous for $K = 0$. Thus, our construction of $\boldsymbol{\xi}^u$ is consistent with all conditions (2).

Let us next consider the one-stage case with a one-dimensional random parameter in the objective function only, that is, $H = 1$, $K = 1$, and $L = 0$. Condition (C4) implies that the given marginal distribution of $\boldsymbol{\eta}$ is supported on a compact convex set $\Theta \subset \mathbb{R}$. By a slight abuse of notation, the two extreme points of Θ are again denoted by a and b, see Figure 2. Next, we construct the distribution of $\boldsymbol{\eta}^u$ conditional on $\boldsymbol{\eta}$. It will be convenient to let this conditional distribution be supported on the single point $\mathrm{E}(\boldsymbol{\eta})$ with conditional probability equal to 1 for all realizations of $\boldsymbol{\eta}$. Thus, the marginal distribution of $\boldsymbol{\eta}^u$ is the Dirac distribution concentrated at $\mathrm{E}(\boldsymbol{\eta})$, see Figure 2. By construction, we find

$$\mathrm{E}(\boldsymbol{\eta}|\mathcal{F}^u) = \mathrm{E}(\boldsymbol{\eta}) = \boldsymbol{\eta}^u \quad P\text{-a. s.},$$

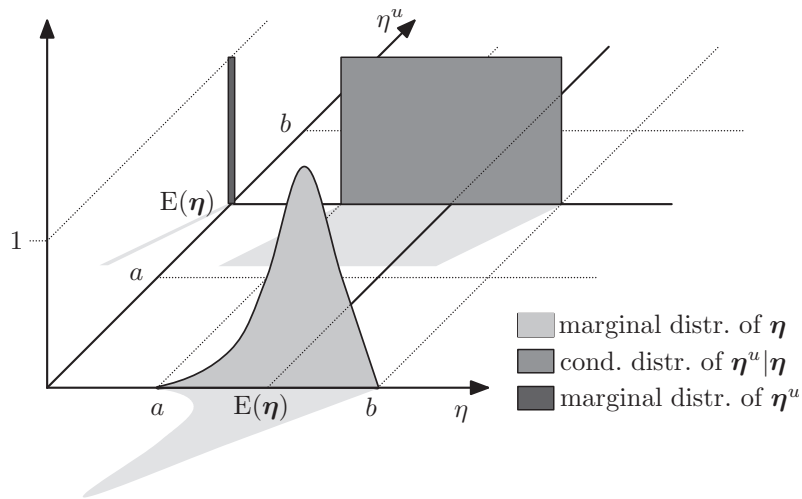

Fig. 2. Relation between $\boldsymbol{\eta}$ and $\boldsymbol{\eta}^u$.

which implies (2d). The conditions (2a) and (2b) are automatically satisfied in the one-stage case, and (2c) becomes vacuous for $L = 0$. Thus, our construction of $\boldsymbol{\eta}^u$ is consistent with all conditions (2).

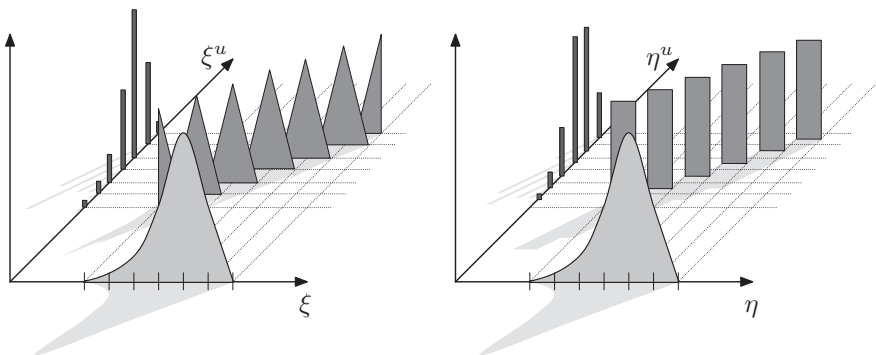

Fig. 3. Refinement of discretization.

It is apparent from Figures 1 and 2 that the marginal distributions of $\boldsymbol{\xi}^u$ and $\boldsymbol{\eta}^u$ can differ substantially from the marginal distributions of $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$, respectively. Hence, the solutions of the original and approximate stochastic programs are likely to differ as well. In order to improve the accuracy of the approximate solution, we refine the construction of the conditional distributions. This can be achieved by applying the basic construction on increasingly small subintervals of Ξ and Θ, respectively.

In the special case $H = L = 1$ and $K = 0$ with $\Xi = [a, b]$, for instance, we select

an increasing sequence of real numbers $a = a_0 < a_1 < \cdots < a_J = b$ and set

$$P(\boldsymbol{\xi}^u = a_j \,|\, \boldsymbol{\xi} = \xi) = \begin{cases} \frac{\xi - a_{j-1}}{a_j - a_{j-1}} & \text{for } a_{j-1} \leq \xi < a_j, \\ \frac{a_j - \xi}{a_{j+1} - a_j} & \text{for } a_j \leq \xi < a_{j+1}, \\ 0 & \text{otherwise,} \end{cases}$$

for $j = 0, \ldots, J$. As opposed to a simple two-point distribution, the marginal distribution of $\boldsymbol{\xi}^u$ now exhibits $J+1$ discretization points. Thus, it bears more resemblance to the marginal distribution of $\boldsymbol{\xi}$, see the left panel of Figure 3. By using familiar arguments, we can verify that the conditions (2) are still satisfied. Moreover, if the diameters of all subintervals $[a_{j-1}, a_j]$ become uniformly small, then the $\mathcal{L}^\infty$-distance between $\boldsymbol{\xi}^u$ and $\boldsymbol{\xi}$ tends to zero.

In the special case $H = K = 1$ and $L = 0$ with $\Theta = [a, b]$, we also select real numbers $a = a_0 < a_1 < \cdots < a_J = b$ and define the conditional distribution of $\boldsymbol{\eta}^u$ through

$$P(\boldsymbol{\eta}^u = \mu_j \,|\, \boldsymbol{\xi} = \xi) = \begin{cases} 1 & \text{for } a_{j-1} \leq \xi < a_j, \\ 0 & \text{otherwise,} \end{cases}$$

where μ_i denotes the conditional expectation of $\boldsymbol{\eta}$ within the interval $[a_{j-1}, a_j)$, see the right panel of Figure 3. Again, the validity of (2) is preserved, while the shape of the marginal distribution of $\boldsymbol{\eta}^u$ is improved. As the diameters of all subintervals $[a_{j-1}, a_j)$ become uniformly small, the $\mathcal{L}^\infty$-distance between $\boldsymbol{\eta}^u$ and $\boldsymbol{\eta}$ converges to zero.

We will not give details on how to construct the conditional distribution of $(\boldsymbol{\eta}^u, \boldsymbol{\xi}^u)$ in the general case $K, L \geq 0$. A complete description is provided in [21, § 4]. We merely remark that the compact convex sets $\Theta \subset \mathbb{R}^K$ and $\Xi \subset \mathbb{R}^L$ are covered by nondegenerate simplices. Moreover, the conditional distribution of $(\boldsymbol{\eta}^u, \boldsymbol{\xi}^u)$ given $(\boldsymbol{\eta}, \boldsymbol{\xi}) = (\eta, \xi)$ is concentrated to some generalized barycenters of the simplices covering Θ and to the vertices of the simplices covering Ξ. The corresponding probabilities are piecewise constant in η and piecewise linear in ξ.

In the multistage case $H \geq 1$, the relations (2a) and (2b) become restrictive. In fact, the inclusion (2a) essentially means that the sets of random vectors $\{\boldsymbol{\eta}_i^u, \boldsymbol{\xi}_i^u\}_{i \leq h}$ and $\{\boldsymbol{\eta}_i, \boldsymbol{\xi}_i\}_{i > h}$ are conditionally independent given $\{\boldsymbol{\eta}_i, \boldsymbol{\xi}_i\}_{i \leq h}$ for all stage indices h. Conditional independence, in turn, follows from the serial independence of $(\boldsymbol{\eta}, \boldsymbol{\xi})$ if $(\boldsymbol{\eta}^u, \boldsymbol{\xi}^u)$ is constructed by using a 'non-anticipative' forward recursion scheme. The inclusion (2b), on the other hand, is essentially equivalent to conditional independence of $\{\boldsymbol{\eta}_i, \boldsymbol{\xi}_i\}_{i \leq h}$ and $\{\boldsymbol{\eta}_i^u, \boldsymbol{\xi}_i^u\}_{i > h}$ given $\{\boldsymbol{\eta}_i^u, \boldsymbol{\xi}_i^u\}_{i \leq h}$ for all h. Since the original $\boldsymbol{\xi}$ process is serially independent, the generalized Jensen inequality (3) is satisfied if the approximate process $\boldsymbol{\xi}^u$ is constructed in a non-anticipative manner; a formal proof is provided in [23, § 5].

In summary, we argue that under the regularity conditions (C1) – (C4) it is possible to systematically construct approximate data processes $\boldsymbol{\eta}^u$ and $\boldsymbol{\xi}^u$ which satisfy both (2) and (3). At the same time, the $\mathcal{L}^\infty$-distance between $(\boldsymbol{\eta}^u, \boldsymbol{\xi}^u)$ and $(\boldsymbol{\eta}, \boldsymbol{\xi})$ can be made as small as desired. Furthermore, it can be shown that the marginal

distribution of $(\boldsymbol{\eta}^u, \boldsymbol{\xi}^u)$ represents a 'barycentric scenario tree' in the sense of Frauendorfer [9, 10]. A major benefit of the novel approach presented here originates from the fact that we do not consider this scenario tree in isolation. Instead, the original and approximate data processes are constructed as dependent random vectors on a joint probability space. In the next section we will exploit this probabilistic coupling to convert optimal solutions for the approximate problem $\mathcal{P}^u$ to implementable and near-optimal decision rules for the original problem $\mathcal{P}$.

Before we can elaborate our main approximation result, we introduce a second pair of discrete stochastic processes $\boldsymbol{\eta}^l = (\boldsymbol{\eta}^l_1, \ldots, \boldsymbol{\eta}^l_H)$ and $\boldsymbol{\xi}^l = (\boldsymbol{\xi}^l_1, \ldots, \boldsymbol{\xi}^l_H)$ which are also defined as measurable mappings from (Ω, Σ, P) to $\mathbb{R}^{HK}$ and $\mathbb{R}^{HL}$, respectively. As usual, we denote by $\mathbb{F}^l = \{\mathcal{F}^l_h\}_{h=1}^H$ the filtration generated by $(\boldsymbol{\eta}^l, \boldsymbol{\xi}^l)$ and set $\mathcal{F}^l = \mathcal{F}^l_H$. It is possible to construct these approximate processes in such a way that the following conditions are satisfied for some versions of the conditional expectations, respectively.

$$\mathrm{E}(\boldsymbol{x}|\mathcal{F}^l) \in X(\mathbb{F}^l) \text{ for all } \boldsymbol{x} \in X(\mathbb{F}) \tag{4a}$$

$$\mathrm{E}(\boldsymbol{y}|\mathcal{F}) \in Y(\mathbb{F}) \text{ for all } \boldsymbol{y} \in Y(\mathbb{F}^l) \tag{4b}$$

$$\mathrm{E}(\boldsymbol{\xi}|\mathcal{F}^l) = \boldsymbol{\xi}^l \tag{4c}$$

$$\mathrm{E}(\boldsymbol{\eta}^l|\mathcal{F}) = \boldsymbol{\eta} \tag{4d}$$

Here, the function spaces $X(\mathbb{F}^l)$ and $Y(\mathbb{F}^l)$ are defined in the obvious manner. Furthermore, we may require the following inequality to hold for $h = 1, \ldots, H$.

$$\mathrm{E}(f_h(\boldsymbol{x}, \boldsymbol{\xi})|\mathcal{F}^l) \geq f_h(\mathrm{E}(\boldsymbol{x}|\mathcal{F}^l), \boldsymbol{\xi}^l) \quad P\text{-a. s.} \quad \text{for all} \quad \boldsymbol{x} \in X(\mathbb{F}) \tag{5}$$

The construction of $\boldsymbol{\eta}^l$ and $\boldsymbol{\xi}^l$ proceeds in complete analogy to the construction of $\boldsymbol{\eta}^u$ and $\boldsymbol{\xi}^u$. The only difference is that the roles of $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are interchanged.

4. APPROXIMATION

In the remainder of this paper we assume that we are given a pair of discrete data processes $\boldsymbol{\eta}^u$ and $\boldsymbol{\xi}^u$ subject to (2) and (3). We further assume that there are discrete processes $\boldsymbol{\eta}^l$ and $\boldsymbol{\xi}^l$ satisfying the conditions (4) and (5). A new approximate stochastic program $\mathcal{P}^l$ is defined analogously to problem $\mathcal{P}^u$: it is obtained from the original problem $\mathcal{P}$ by replacing $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ with $\boldsymbol{\eta}^l$ and $\boldsymbol{\xi}^l$, respectively. We also substitute $\mathbb{F}^l$ for the original filtration $\mathbb{F}$. With these preparations, we now introduce our central approximation result.[3]

Theorem 2. (Kuhn et al. [23, Theorem 5.5]) Assume that the approximate problems $\mathcal{P}^l$ and $\mathcal{P}^u$ are solvable with finite optimal values. If $\boldsymbol{x}^u$ solves $\mathcal{P}^u$, then $\hat{\boldsymbol{x}} = \mathrm{E}(\boldsymbol{x}^u|\mathcal{F})$ is feasible in $\mathcal{P}$ and

$$\inf \mathcal{P}^l \leq \inf \mathcal{P} \leq \mathrm{E}(c(\hat{\boldsymbol{x}}, \boldsymbol{\eta})) \leq \inf \mathcal{P}^u.$$

[3]Theorem 2 is inspired by Theorem 2 in Birge and Louveaux [2, § 11.1].

Proof. We repeat the proof from [23] to keep this paper self-contained and to illustrate the intuition behind the assumptions (2) – (5). Using Lemma 1 to reformulate $\mathcal{P}$ as an unconstrained min-max problem, we find

$$\begin{aligned}
\inf \mathcal{P} &\geq \inf_{\boldsymbol{x}\in X(\mathbb{F})} \sup_{\boldsymbol{y}\in Y(\mathbb{F}^l)} \mathrm{E}\left(c(\boldsymbol{x}, \mathrm{E}(\boldsymbol{\eta}^l|\mathcal{F})) + \sum_{h=1}^{H} \mathrm{E}(\boldsymbol{y}_h|\mathcal{F}) \cdot f_h(\boldsymbol{x}, \boldsymbol{\xi}) \right) \\
&\geq \inf_{\boldsymbol{x}\in X(\mathbb{F})} \sup_{\boldsymbol{y}\in Y(\mathbb{F}^l)} \mathrm{E}\left(c(\boldsymbol{x}, \boldsymbol{\eta}^l) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot f_h(\boldsymbol{x}, \boldsymbol{\xi}) \right) \\
&= \inf_{\boldsymbol{x}\in X(\mathbb{F})} \sup_{\boldsymbol{y}\in Y(\mathbb{F}^l)} \mathrm{E}\left(\mathrm{E}(c(\boldsymbol{x}, \boldsymbol{\eta}^l)|\mathcal{F}^l) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot \mathrm{E}(f_h(\boldsymbol{x}, \boldsymbol{\xi})|\mathcal{F}^l) \right).
\end{aligned}$$

Here, the first inequality is a direct consequence of (4b) and (4d). The second inequality follows from the conditional Jensen inequality, which applies since the cost function is concave in its second argument, while $\boldsymbol{x}$ and $\boldsymbol{\xi}$ are $\mathcal{F}$-measurable. Finally, the equality in the third line follows from the law of iterated conditional expectations and uses the fact that $\boldsymbol{y}$ is $\mathcal{F}^l$-measurable. Next, we have

$$\begin{aligned}
\inf \mathcal{P} &\geq \inf_{\boldsymbol{x}\in X(\mathbb{F})} \sup_{\boldsymbol{y}\in Y(\mathbb{F}^l)} \mathrm{E}\left(c(\mathrm{E}(\boldsymbol{x}|\mathcal{F}^l), \boldsymbol{\eta}^l) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot f_h(\mathrm{E}(\boldsymbol{x}|\mathcal{F}^l), \boldsymbol{\xi}^l) \right) \\
&\geq \inf_{\boldsymbol{x}\in X(\mathbb{F}^l)} \sup_{\boldsymbol{y}\in Y(\mathbb{F}^l)} \mathrm{E}\left(c(\boldsymbol{x}, \boldsymbol{\eta}^l) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot f_h(\boldsymbol{x}, \boldsymbol{\xi}^l) \right),
\end{aligned}$$

where the first inequality holds by (5) and the conditional Jensen inequality. Note that the cost function is convex in its first argument and that the approximate process $\boldsymbol{\eta}^l$ is $\mathcal{F}^l$-measurable. The second inequality is due to (4a). So far, we have therefore shown $\inf \mathcal{P}^l \leq \inf \mathcal{P}$.

Observe now that $\boldsymbol{x}^u \in X(\mathbb{F}^u)$, which implies via (2a) that $\mathrm{E}(\boldsymbol{x}^u|\mathcal{F}) \in X(\mathbb{F})$. This justifies the first inequality in

$$\begin{aligned}
\inf \mathcal{P} &\leq \sup_{\boldsymbol{y}\in Y(\mathbb{F})} \mathrm{E}\left(c(\mathrm{E}(\boldsymbol{x}^u|\mathcal{F}), \boldsymbol{\eta}) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot f_h(\mathrm{E}(\boldsymbol{x}^u|\mathcal{F}), \boldsymbol{\xi}) \right) \\
&\leq \sup_{\boldsymbol{y}\in Y(\mathbb{F})} \mathrm{E}\left(\mathrm{E}(c(\boldsymbol{x}^u, \boldsymbol{\eta})|\mathcal{F}) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot \mathrm{E}(f_h(\boldsymbol{x}^u, \boldsymbol{\xi}^u)|\mathcal{F}) \right) \qquad (6) \\
&= \sup_{\boldsymbol{y}\in Y(\mathbb{F})} \mathrm{E}\left(c(\boldsymbol{x}^u, \boldsymbol{\eta}) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot f_h(\boldsymbol{x}^u, \boldsymbol{\xi}^u) \right).
\end{aligned}$$

The second inequality is based on (3) and the conditional Jensen inequality, while the equalityuses the law of iterated conditional expectations and the fact that $\boldsymbol{y}$ is

$\mathcal{F}$-measurable. Another application of the conditional Jensen inequality then yields

$$\begin{aligned}
\inf \mathcal{P} &\leq \sup_{\boldsymbol{y} \in Y(\mathbb{F})} \mathrm{E}\left(c(\boldsymbol{x}^u, \mathrm{E}(\boldsymbol{\eta}|\mathcal{F}^u)) + \sum_{h=1}^{H} \mathrm{E}(\boldsymbol{y}_h|\mathcal{F}^u) \cdot f_h(\boldsymbol{x}^u, \boldsymbol{\xi}^u) \right) \\
&\leq \sup_{\boldsymbol{y} \in Y(\mathbb{F}^u)} \mathrm{E}\left(c(\boldsymbol{x}^u, \boldsymbol{\eta}^u) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot f_h(\boldsymbol{x}^u, \boldsymbol{\xi}^u) \right) \\
&= \inf_{\boldsymbol{x} \in X(\mathbb{F}^u)} \sup_{\boldsymbol{y} \in Y(\mathbb{F}^u)} \mathrm{E}\left(c(\boldsymbol{x}, \boldsymbol{\eta}^u) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot f_h(\boldsymbol{x}, \boldsymbol{\xi}^u) \right).
\end{aligned}$$

The second inequality holds by (2b) and (2d), which entails a relaxation of the dual feasible set. The last line in the above expression can easily be identified as $\inf \mathcal{P}^u$, which is finite by assumption. Hence, the supremum over $Y(\mathbb{F})$ in the first line of (6) is finite, too, and it is attained by the trivial strategy $\boldsymbol{y} = \mathbf{0}$. The primal strategy $\hat{\boldsymbol{x}} = \mathrm{E}(\boldsymbol{x}^u|\mathcal{F})$ is therefore feasible in $\mathcal{P}$, and its objective value $\mathrm{E}(c(\hat{\boldsymbol{x}}, \boldsymbol{\eta}))$ satisfies the postulated inequalities. □

Theorem 2 addresses the following fundamental problem in stochastic programming: if the support of $(\boldsymbol{\eta}, \boldsymbol{\xi})$ is infinite (or of finite but large cardinality), then the original stochastic program $\mathcal{P}$ is computationally untractable (or extremely hard to solve). In such situations, Theorem 2 provides an a priori estimate for the minimal expected cost achievable in reality, that is, the optimal value of problem $\mathcal{P}$. This estimate is expressed in terms of upper and lower bounds, which are typically computable since they are given by the optimal values of two finite-dimensional convex optimization problems.[4] Notice that the solutions of $\mathcal{P}^l$ and $\mathcal{P}^u$ depend only on the marginal distributions of $(\boldsymbol{\eta}^l, \boldsymbol{\xi}^l)$ and $(\boldsymbol{\eta}^u, \boldsymbol{\xi}^u)$, respectively. The relations between these processes and the original data process, on the other hand, do not affect the approximate solutions.

A typical decision maker is not only interested in estimating the optimal value of $\mathcal{P}$, which is attained by some unknown ideal strategy, but even more in finding a computable strategy whose objective value comes close to the theoretical optimum. A naive guess would be that the optimal solutions of $\mathcal{P}^l$ and or $\mathcal{P}^u$ might constitute near-optimal solutions for $\mathcal{P}$. However, these strategies are not even feasible in $\mathcal{P}$ since they are confined to a relatively small scenario set (that is, the support of the approximate data processes). Theorem 2 resolves this dilemma by proposing a policy $\hat{\boldsymbol{x}}$ which is implementable for all scenarios (that is, the support of the original data processes) and whose expected cost is bracketed by $\inf \mathcal{P}^l$ and $\inf \mathcal{P}^u$. Here, we exploit the stochastic interdependence of $(\boldsymbol{\eta}, \boldsymbol{\xi})$ and $(\boldsymbol{\eta}^u, \boldsymbol{\xi}^u)$ to transform the optimal solution $\boldsymbol{x}^u$ of $\mathcal{P}^u$ to a feasible near-optimal solution $\hat{\boldsymbol{x}}$ of $\mathcal{P}$. Note that $\mathrm{E}(c(\hat{\boldsymbol{x}}, \boldsymbol{\eta}))$ represents the expected cost which is achieved in reality by implementing $\hat{\boldsymbol{x}}$.

Since $\hat{\boldsymbol{x}}$ constitutes an $\mathcal{F}$-measurable random vector, there is a Borel measurable decision rule $\hat{\rho} : \mathbb{R}^{H(K+L)} \to \mathbb{R}^{HN}$ such that $\hat{\boldsymbol{x}} = \hat{\rho}(\boldsymbol{\eta}, \boldsymbol{\xi})$. The decision rule $\hat{\rho}$ assigns

[4] In applications, the cost and constraint functions are frequently linear or convex quadratic. In these cases, and given enough storage space and processing power, the finite-dimensional approximate problems $\mathcal{P}^l$ and $\mathcal{P}^u$ can be solved by linear or quadratic programming algorithms.

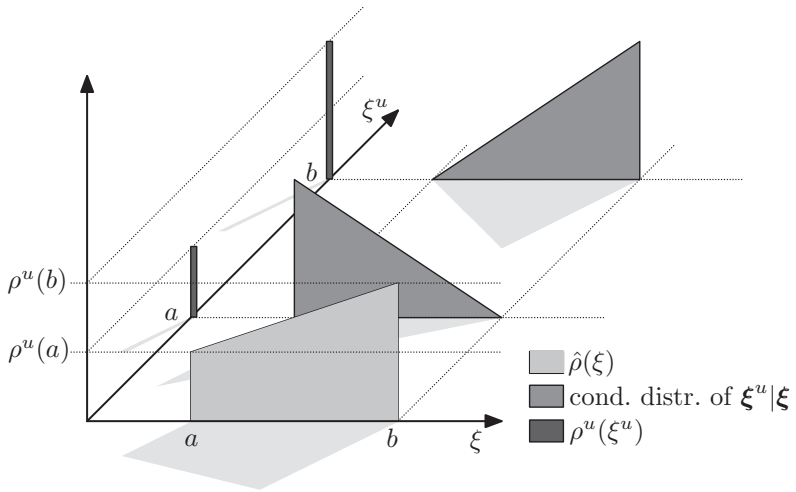

Fig. 4. Decision rules.

to each scenario $(\eta, \xi) \in \mathbb{R}^{H(K+L)}$ a sequence of H actions in $\mathbb{R}^N$. Similarly, since $\boldsymbol{x}^u$ constitutes an $\mathcal{F}^u$-measurable random vector, there is a Borel measurable decision rule ρ^u which has the same domain and range as $\hat{\rho}$ and satisfies $\boldsymbol{x}^u = \rho^u(\boldsymbol{\eta}^u, \boldsymbol{\xi}^u)$. Note that only the restriction of ρ^u to the support of the approximate data processes has physical meaning. Moreover, the value of ρ^u for any realization of $(\boldsymbol{\eta}^u, \boldsymbol{\xi}^u)$ can easily be read off from the solution $\boldsymbol{x}^u$ of $\mathcal{P}^u$. The implementable decision rule $\hat{\rho}$, on the other hand, is only given implicitly in terms of a conditional expectation of $\boldsymbol{x}^u$. Since $\boldsymbol{x}^u$ is finitely supported, evaluation of $\hat{\rho}(\eta, \xi)$ reduces to the evaluation of a finite sum, which is computationally cheap. For the further argumentation, we let $\{(\eta_j^u, \xi_j^u)\}_{j=1}^J$ be the finite support of $(\boldsymbol{\eta}^u, \boldsymbol{\xi}^u)$. Furthermore, we introduce conditional probabilities

$$p_j^u(\eta, \xi) = P\left((\boldsymbol{\eta}^u, \boldsymbol{\xi}^u) = (\eta_j^u, \xi_j^u) \,|\, (\boldsymbol{\eta}, \boldsymbol{\xi}) = (\eta, \xi)\right)$$

for all scenarios $j = 1, \ldots, J$, where (η, ξ) is any point in the support of $(\boldsymbol{\eta}, \boldsymbol{\xi})$. Observe that these conditional probabilities are known. In fact, we prescribed them ourselves when specifying the approximate data processes, see Section 3. The value of $\hat{\rho}$ at point (η, ξ) thus amounts to

$$\hat{\rho}(\eta, \xi) = \sum_{j=1}^{J} p_j^u(\eta, \xi) \rho^u(\eta_j^u, \xi_j^u)\,.$$

Figure 4 illustrates how to obtain $\hat{\rho}$ in the special case $H = L = 1$, $K = 0$, and $\Xi = [a, b]$. To keep this example as simple as possible, we assume that the support of $\boldsymbol{\xi}^u$ contains only the extreme scenarios a and b, which are indexed by 1 and 2, respectively. By construction, $p_1^u(\xi)$ decreases linearly from 1 to 0 as ξ increases from a to b. At the same time, the conditional probability $p_2^u(\xi)$ increases linearly

from 0 to 1, see Section 3. Assume now that the optimal solution of the approximate problem $\mathcal{P}^u$ prescribes the actions $\rho^u(a)$ and $\rho^u(b)$ in scenarios a and b, respectively. Then, the decision rule $\hat{\rho}$ linearly interpolates the optimal decisions $\rho^u(a)$ and $\rho^u(b)$ for all $\xi \in [a, b]$.

It is easy to visualize that the decision rule $\hat{\rho}$ will be piecewise linear if the discretization of $\boldsymbol{\xi}$ is refined as in the left panel of Figure 3. One also easily verifies that the decision rule $\hat{\rho}$ will turn out to be piecewise constant in the special case $H = K = 1$ and $L = 0$. In the general multistage case with $H \geq 0$ and K, L arbitrary, the conditional probabilities $p_j^u(\eta, \xi)$ are piecewise constant in η and piecewise polynomial in ξ.[5] This property is inherited by the decision rule $\hat{\rho}$.

In conclusion, it is worthwhile to remark that the gap between the upper and lower bounds in Theorem 2 converges to zero as the discretizations of the data processes are suitably refined, see e. g. [22, 23]. At the same time, the expected cost of the implementable strategy $\hat{\boldsymbol{x}}$ converges to the optimal value of the original problem $\mathcal{P}$. Note also that the expected cost of $\hat{\boldsymbol{x}}$ can conveniently be calculated by Monte Carlo simulation.

5. EXAMPLE: MEAN–RISK PORTFOLIO OPTIMIZATION

Expected value constraints are a central element of mean-risk portfolio selection models. Therefore, such models are ideal candidates to illustrate our methodologies. We consider a market of L stocks which can be traded at time points $0 = t_1 < \cdots < t_H = T$.[6] For each h the L-dimensional random vector $\boldsymbol{\xi}_h$ characterize the price relatives of the stocks over the interval from t_{h-1} to t_h. The price relative of a stock over a certain period is given by the ratio of terminal price to initial price. By convention, we denote by $\boldsymbol{\xi} = \{\boldsymbol{\xi}_h\}_{h=1}^H$ the process of price relatives and let $\mathbb{F} = \{\mathcal{F}_h\}_{h=1}^H$ be the filtration generated by $\boldsymbol{\xi}$. Next, we specify the decision variables. We set $\boldsymbol{w}_h^{\pm} = (\boldsymbol{w}_{1,h}^{\pm}, \ldots, \boldsymbol{w}_{L,h}^{\pm})$, where $\boldsymbol{w}_{l,h}^-$ denotes the capital invested in asset l at time t_h before portfolio rebalancing, while $\boldsymbol{w}_{l,h}^+$ denotes the capital invested in asset l at time t_h after reallocation of funds. Moreover, we set $\boldsymbol{b}_h = (\boldsymbol{b}_{1,h}, \ldots, \boldsymbol{b}_{L,H})$ and $\boldsymbol{s}_h = (\boldsymbol{s}_{1,h}, \ldots, \boldsymbol{s}_{L,H})$, where $\boldsymbol{b}_{l,h}$ and $\boldsymbol{s}_{l,h}$ represent the amounts of money used at time t_h to buy and sell stocks of type l, respectively. With these conventions, we can now formulate the problem of minimizing portfolio risk subject to a performance constraint.

$$
\begin{array}{lll}
\multicolumn{3}{c}{\text{minimize } \mathrm{CVaR}_\alpha(e^\top \boldsymbol{w}_1^- - e^\top \boldsymbol{w}_H^-)} \\
\text{s.t.} & \mathrm{E}(e^\top \boldsymbol{w}_H^-) \geq \gamma\, e^\top \boldsymbol{w}_1^- & \\
& \boldsymbol{w}_h^+ = \boldsymbol{w}_h^- + \boldsymbol{b}_h - \boldsymbol{s}_h & 1 \leq h < H \\
& (1 + c_b)\, e^\top \boldsymbol{b}_h = (1 - c_s)\, e^\top \boldsymbol{s}_h & \text{"} \\
& \boldsymbol{w}_{h+1}^- = \boldsymbol{\xi}_{h+1} \odot \boldsymbol{w}_h^+ & \text{"} \\
& \boldsymbol{w}_h^-, \boldsymbol{w}_h^+, \boldsymbol{b}_h, \boldsymbol{s}_h \geq 0,\ \mathcal{F}_h\text{-measurable} & \text{"}
\end{array}
\qquad (\mathcal{P}')
$$

[5] In the multistage case, the $p_j^u(\eta, \xi)$ are products of H stagewise conditional probabilities, which are all piecewise constant in η and piecewise linear in ξ, see [21, § 4].

[6] For notational convenience, will sometimes use an additional time point $t_0 = 0$.

The parameter γ represents the return target over the investment period, c_b and c_s denote the transaction costs per unit of currency for sales and purchases of the stocks, respectively, and '$\odot$' denotes the entrywise Hadamard product. Moreover, the constant vector $\boldsymbol{w}_1^-$ characterizes the initial portfolio. To measure risk, we use the conditional value-at-risk (CVaR) at level $\alpha \in (0,1)$. Rockafellar and Uryasev [28] have shown that if $\boldsymbol{L}$ is some random variable representing 'loss', then

$$\mathrm{CVaR}_\alpha(\boldsymbol{L}) = \inf_{\beta\in\mathbb{R}} \left\{\beta + (1-\alpha)^{-1}\,\mathrm{E}(\boldsymbol{L}-\beta)_+\right\}. \tag{7}$$

In our case, loss corresponds to the difference between initial and terminal portfolio value. By using the representation (7), problem $\mathcal{P}'$ can be brought to the standard form $\mathcal{P}$. One can easily verify that this problem satisfies the regularity conditions (C1) – (C4) if the price relatives are bounded and serially independent.

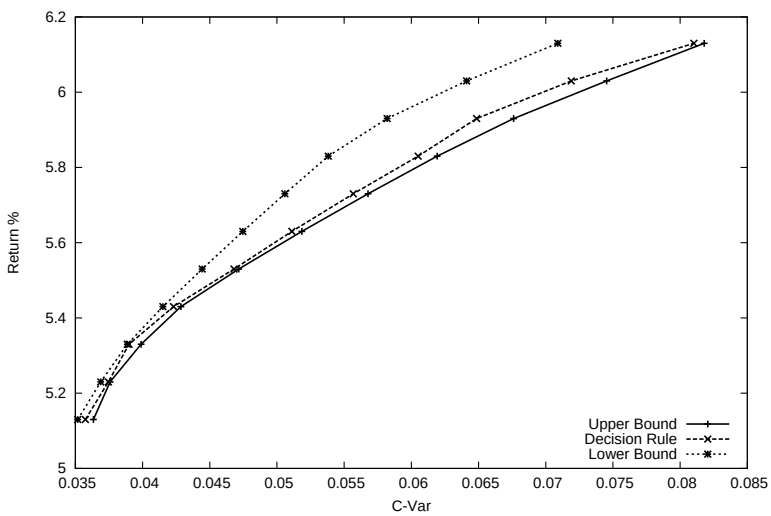

Fig. 5. Approximate efficient frontiers.

Our computational example is based on a market model with $L = 5$ assets and $H = 4$ monthly investment intervals. The stock returns $\{\boldsymbol{\xi}_h\}_{h=1}^H$ are assumed to be independent and identically distributed. In particular, the monthly log-returns follow a truncated multivariate normal distribution with mean values 0.41 %, 0.82 %, 1.19 %, 1.60 %, 1.99 % and standard deviations 8.5 %, 8.0 %, 9.5 %, 9.0 %, 10.0 %. The correlations are set to 30% uniformly over all pairs of stocks. We aim at solving mean-risk portfolio problems of the type $\mathcal{P}'$ with $\alpha = 95\%$ and for different values of the return target parameter γ. We start from a regular reformulation $\mathcal{P}$ of $\mathcal{P}'$ and construct upper and lower bounding problems $\mathcal{P}^u$ and $\mathcal{P}^l$, respectively. The construction of these bounding problems goes along the lines of Sections 3 and 4. Unlike $\mathcal{P}$, the approximate problems allow for numerical solution. By solving them for different values of γ, we can trace out two closely aligned efficient frontiers in the

risk-return plane, see Figure 5. Theorem 2 guarantees that these upper and lower approximate frontiers bracket the true efficient frontier. Moreover, all portfolios corresponding to decision rules $\hat{\rho}$ in the sense of Section 4 lie also between the approximate frontiers in the risk-return plane. Note that these decision rules are computable and implementable in reality, while the solutions of $\mathcal{P}^u$ and $\mathcal{P}^l$ are not implementable, and $\mathcal{P}$ is not directly solvable. The risk associated with the decision rules was estimated via Monte Carlo simulation. The number of sample paths (5×10^6) was chosen to guarantee that the standard error of the estimator is smaller than 0.001%.

ACKNOWLEDGEMENT

This work was supported by EPSRC grants GR/T02560/01 and EP/C513584/1.

(Received November 5, 2007.)

REFERENCES

[1] R. Ash: Real Analysis and Probability. Probability and Mathematical Statistics. Academic Press, Berlin 1972.
[2] J. Birge and F. Louveaux: Introduction to Stochastic Programming. Springer-Verlag, New York 1997.
[3] J. Birge and R.-B. Wets: Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse. Math. Programming Stud. *27* (1986), 54–102.
[4] S. P. Dokov and D. P. Morton: Second-order lower bounds on the expectation of a convex function. Math. Oper. Res. *30* (2005), 3, 662–677.
[5] J. Dupačová: Stochastic programming: Minimax approach. In: Encyclopedia of Optimization (C. Floudas and P. Pardalos, eds.), vol. 5, Kluwer 2001, pp. 327–330.
[6] J. Dupačová (as Žáčková): On minimax solutions of stochastic linear programming problems. Časopis pro pěstování matematiky *91* (1966), 423–429.
[7] N. Edirisinghe and W. Ziemba: Bounding the expectation of a saddle function with application to stochastic programming. Math. Oper. Res. *19* (1994), 314–340.
[8] S.-E. Fleten, K. Høyland, and S. W. Wallace: The performance of stochastic dynamic and fixed mix portfolio models. European J. Oper. Res. *140* (2002), 1, 37–49.
[9] K. Frauendorfer: Multistage stochastic programming: Error analysis for the convex case. Z. Oper. Res. *39* (1994), 1, 93–122.
[10] K. Frauendorfer: Barycentric scenario trees in convex multistage stochastic programming. Math. Programming *75* (1996), 2, 277–294.
[11] S. J. Garstka and R. J.-B. Wets: On decision rules in stochastic programming. Math. Programming *7* (1974), 117–143.
[12] W. K. Haneveld and M. van der Vlerk: Integrated chance constraints: Reduced forms and an algorithm. Comput. Manag. Sci. *3* (2006), 2, 245–269.
[13] H. Heitsch, W. Römisch, and C. Strugarek: Stability of multistage stochastic programs. SIAM J. Optim. *17* (2006), 511–525.
[14] R. Hochreiter and G. Pflug,: Financial scenario generation for stochastic multi-stage decision processes as facility location problems. Ann. Oper. Res. *156* (2007), 1, 257–272.
[15] K. Høyland and S. Wallace: Generating scenario trees for multistage decision problems. Management Sci. *47* (2001), 2, 295–307.

[16] G. Infanger: Planning under Uncertainty: Solving Large-Scale Stochastic Linear Programs. Boyd and Fraser, Danvers 1994.
[17] V. Kaňková and M. Šmíd: On approximation in multistage stochastic programs: Markov dependence. Kybernetika *40* (2004), 5, 625–638.
[18] A. J. Kleywegt, A. Shapiro, and T. Homem-de-Mello: The sample average approximation method for stochastic discrete optimization. SIAM J. Optim. *12* (2002), 2, 479–502.
[19] M. Koivu: Variance reduction in sample approximations of stochastic programs. Math. Programming, Ser. A *103* (2005), 3, 463–485.
[20] R. Kouwenberg: Scenario generation and stochastic programming models for asset and liability management. European J. Oper. Res. *134* (2001), 2, 279–292.
[21] D. Kuhn: Aggregation and discretization in multistage stochastic programming. Math. Programming, Ser. A *113* (2008), 1, 61–94.
[22] D. Kuhn: Convergent bounds for stochastic programs with expected value constraints. The Stochastic Programming E-Print Series (SPEPS), 2007.
[23] D. Kuhn, P. Parpas, and B. Rustem: Threshold accepting approach to improve bound-based approximations for portfolio optimization. In: Computational Methods in Financial Engineering (E. Kontoghiorghes, B. Rustem, and P. Winker, eds.), Springer–Verlag, Berlin 2008, pp. 3–26.
[24] R. Mirkov and G. Pflug: Tree approximations of dynamic stochastic programs. SIAM J. Optim. *18* (2007), 3, 1082–1105.
[25] T. Pennanen: Epi-convergent discretizations of multistage stochastic programs via integration quadratures. Math. Programming, to appear.
[26] G. Pflug: Scenario tree generation for multiperiod financial optimization by optimal discretization. Math. Programming, Ser. B *89* (2001), 251–271.
[27] S. Rachev and W. Römisch: Quantitative stability in stochastic programming: the method of probability metrics. Math. Oper. Res. *27* (2002), 792–818.
[28] R. T. Rockafellar and S. Uryasev: Optimization of conditional value-at-risk. J. Risk *2* (2000) 3, 21–41.
[29] A. Shapiro and A. Nemirovski: On complexity of stochastic programming problems. In: Continuous Optimization: Current Trends and Applications 2005 (V. Jeyakumar and A. Rubinov, eds.), Springer-Verlag, Berlin 2006, pp. 111–144.
[30] J. Thénié and J.-P. Vial: Step decision rules for multistage stochastic programming: a heuristic approach. Optimization Online, 2006.
[31] S. Wright: Primal-dual aggregation and disaggregation for stochastic linear programs. Math. Oper. Res. *19* (1994), 4, 893–908.

Daniel Kuhn, Panos Parpas, and Berç Rustem, Department of Computing, Imperial College of Science, Technology and Medicine, 180 Queen's Gate, London SW7 2BZ. United Kingdom.
e-mails: dkuhn@doc.ic.ac.uk, pp500@doc.ic.ac.uk, br@doc.ic.ac.uk