GUARANTEED BOUNDS FOR GENERAL NONDISCRETE MULTISTAGE RISK-AVERSE STOCHASTIC OPTIMIZATION PROGRAMS[*]

FRANCESCA MAGGIONI[†] AND GEORG CH. PFLUG[‡]

Abstract. In general, multistage stochastic optimization problems are formulated on the basis of continuous distributions describing the uncertainty. Such "infinite" problems are practically impossible to solve as they are formulated, and finite tree approximations of the underlying stochastic processes are used as proxies. In this paper, we demonstrate how one can find guaranteed bounds, i.e., finite tree models, for which the optimal values give upper and lower bounds for the optimal value of the original infinite problem. Typically, there is a gap between the two bounds. However, this gap can be made arbitrarily small by making the approximating trees bushier. We consider approximations in the first-order stochastic sense, in the convex-order sense, and based on subgradient approximations. Their use is shown in a multistage risk-averse production problem.

Key words. bounds, barycentric approximations, first-order stochastic dominance, convex stochastic dominance, risk measures, multistage stochastic programs

AMS subject classifications. 90C15, 90C90, 65K05

DOI. 10.1137/17M1140601

1. Introduction. In this paper, we consider a multistage stochastic optimization problem of the form

$$(1.1)\quad v(P) := \min_{x_0,\ldots,x_T} \{\mathbb{E}[Q(x_0,\xi_1,x_1,\ldots,\xi_T,x_T)] : x_t \lhd \mathcal{F}_t = \sigma(\xi_1,\ldots,\xi_t),\ x_t \in \mathbb{X}_t\},$$

where $Q(\cdot)$ is some cost function, $\xi = (\xi_1,\ldots,\xi_T)$ is the stochastic scenario process defined on a probability space $(\Xi, \mathcal{F}, P)$, where $\Xi = \mathsf{X}_{i=1}^T \Xi_i$ and $\mathcal{F} = (\mathcal{F}_1,\ldots,\mathcal{F}_T)$ is the filtration generated by projections of Ξ onto $\mathsf{X}_{i=1}^t \Xi_i$ for each t. The decision process is $x = (x_0,\ldots,x_T)$ and the notation $x_t \lhd \mathcal{F}_t$ means that x_t is measurable w.r.t. to $\mathcal{F}_t$. This constraint is called the nonanticipativity constraint.[1] $\mathbb{X}_t$ is the set of constraints for the decision variables x_t at stage $t = 1,\ldots,T$, which may be incorporated into the objective function Q by adding the convex indicator function

$$\mathbb{I}_{\mathbb{X}_t}(x_t) = \begin{cases} 0 & \text{if } x_t \in \mathbb{X}_t, \\ \infty & \text{otherwise.} \end{cases}$$

It is well known that in the case of a general nondiscrete process ξ, problem (1.1) is unsolvable and some approximations are needed.

The goal of this paper is to show how one can construct finite processes $\underline{\xi} = (\underline{\xi}_1,\ldots,\underline{\xi}_T)$ and $\overline{\xi} = (\overline{\xi}_1,\ldots,\overline{\xi}_T)$ with respective distributions $\underline{P}$ and $\overline{P}$ such that

---

[†]Department of Management, Economics and Quantitative Methods, University of Bergamo Via dei Caniana 2, 24127 Bergamo, Italy (francesca.maggioni@unibg.it).

[‡]Department of Statistics and Operations Research and International Institute for Applied Systems Analysis (IIASA), University of Vienna, Laxenburg, Austria (georg.pflug@univie.ac.at).

[1]For the sake of simplicity, we do not consider the case where the available information—the filtration—is larger than that generated by ξ. In the latter case one has to consider the nested distribution of ξ (see [25]), but for this paper the multivariate distribution of ξ contains all information.

454

both processes $\underline{\xi}$ and $\overline{\xi}$ take only finitely many values and are tree processes.[2] Then, solving the main problem (1.1) for the finite tree probabilities $\underline{P}$ and $\overline{P}$ one gets for the value function the inequality

$$(1.2) \qquad\qquad v(\underline{P}) \leq v(P) \leq v(\overline{P})$$

for certain classes of cost functions Q. Small corrections of the bounds are allowed, as presented in our main Proposition 3.5. Notice that the bounding tree processes incorporate the nonanticipativity constraints by construction.

The procedure of approximation of the process ξ by tree processes is typically known as the *scenario tree generation.* Many are the papers dealing with scenario tree generation and the calculation of the approximation error associated with the bounding approach. In this paper, we deal with *scenario generation under dominance*, since our construction of "upper" and "lower" trees allows one to run the very same optimization program on the two trees to get the bounds as error estimates.

Given the solutions of the problem on the two trees, we are also interested in finding good solutions for the continuous problem. If the gap between the two solutions is not sufficiently small, one may use finer approximations to improve the bounds. It is assumed in this paper that arbitrarily many paths of the stochastic process ξ can be generated. Our bounds are based on the realizations of a large number of sample paths of ξ.

Review of existing literature. One way to obtain valid bounds is to replace the solution of the original big stochastic optimization problem associated with a large discrete scenario tree with an aggregation of repeated solutions of smaller problems. These techniques are used in [2, 27] for two-stage stochastic programs and are addressed in [3, 9, 20, 21, 22, 28, 29] for the multistage ones. In addition, in [16] the author elaborates an approximation scheme that integrates stage-aggregation and discretization through coarsening of sigma-algebras to ensure computational tractability, while providing deterministic error bounds.

An alternative approach is to construct not just approximating scenario trees, but upper and lower bounds for the optimal value of the original continuous problem in terms of the underlying uncertainty. The advantage of this approach is that it can be refined whenever the gap between the bounds is not sufficiently small.

Two-stage programs. For two-stage programs, the problem reduces to finding good finite discretizations of a probability measure. Many authors have treated the so-called generalized moment problem, i.e., to find extremals of sets of probability measures, which are characterized by some but not full information, such as moments or expectations of other functions. Quite generally, Ermoliev, Gaivoronski, and Nedeva [8] have shown that minimizing an expectation among all probabilities fulfilling ℓ inequality constraints leads to an argmin probability with a support on maximally $\ell + 1$ points, i.e., one gets finite scenario models "automatically." Bounds on the expectation of a convex function are developed in [1, 4, 5, 6, 7] for the two-stage case. In particular, in [6] Edirisinghe and Ziemba propose tight upper and lower bounds in the case of multivariate random variables for two-stage stochastic programs using first and cross moments, also considering the case of unbounded domains. Typically, however, the minimizing probability measure that sits on finitely many points depends on the considered convex function. Our Example 2 shows such a situation.

---

[2] A process $\eta = (\eta_1, \ldots, \eta_T)$ is called a tree proces, if the conditional distribution of $(\eta_1, \ldots, \eta_{t-1})$ given η_t is degenerated, i.e., sits on just one path for all η_t and all t.

Multistage programs. In [5] Edirisinghe applies two-stage bounds-based distributional approximations (i.e., moment-based approximations derived as solutions to certain generalized moment problems) to a multistage stochastic linear program, relaxing nonanticipativity constraints; however, nonanticipativity is regained progressively via a disaggregation procedure. In [7] Edirisinghe and Ziemba propose tight upper and lower bounds to stochastic convex programs in the specific case of random right-hand sides. Using a constraint aggregation procedure, a group of stages from the end of the multistage stochastic program are aggregated to form a single stage and error bounds on the usual mean model are developed.

On the other hand, barycentric bounds that entail the construction of lower and upper scenario trees were proposed in [10, 11] and also [12, 15]. These authors observed that for convex minimization problems in both the random parameters and decision variables, one has to concentrate the probability mass on the *barycenters* of the partition covering the support of the distribution (a discretization of the underlying probability space) to get a lower bound. This is justified by Jensen's inequality. For concave functions in the random parameters, one has to distribute the probability mass to the extreme points to get a lower bound, which is a generalization of the Edmundson–Madansky inequality. Let us call these methods *balayage*.[3] It means that the probability mass is swept as if by a broom (in French, *balai*) either in the center of the domain or at its corners. Such an approach (without naming it balayage) appears in [11, 12]. These authors considered a mixed type of objective, which is concave in one random variable, say η, and convex in another one, say ξ, and assumed that they follow a block-diagonal auto-regressive process. For upper bounds, one has to use the balayage measure for the convex part and the barycentric measure for the concave part. In contrast, the barycentric measure for the convex part and the balayage measure for the concave part provide a lower bound. The bounds can be made arbitrarily tight by successively partitioning the domain of the random vectors, typically by simplices. In [12] the authors discretized the conditional distributions and in a second step the transition probabilities were combined to form a discrete scenario tree. We demonstrate in Examples 5 and 6 that if the conditions required in [11, 12] on the random process are not satisfied, this leads to an approximation but not to a valid bound. In [15] barycentric discretizations were adopted in a more general setting investigating convex multistage stochastic programs with a generalized nonconvex dependence on the random variables.

The novelty of our approach. In this paper, we generalize the well-known bounding ideas of the balayage [10, 11, 12, 15] to not necessarily Markovian scenario processes and derive valid lower and upper bounds for the monotonic and convex cases in both the random parameters and decision variables.

We construct new discrete trees directly from the simulated data of the whole scenario process and not from the discretization of the conditional distributions as done before in the literature, e.g., in [7] and [10], by means of the concept of stochastic dominance of probability measures. We show in Example 5 that it is not sufficient to use a stepwise procedure of discretizing first the distribution of ξ_1 and then the conditional distributions ξ_2 given the previously determined finitely many values for ξ_1 to get a valid bound.

By constructing upper and lower trees, we take care of the fact that the scenario tree approximation keeps the nonanticipativity requirements of the original problem.

---

[3] The word "balayage" was introduced by Choquet; see, for instance, [23].

Our bounds are the optimal values of the problem at hand on finite scenario trees and therefore are easy to get without changing the optimization model. Moreover, we relate the solution of the approximating tree to the solution of the original continuous problem and vice versa; that is by solving the upper and lower tree approximations, we automatically get ϵ-solutions, i.e., near optimal solutions for the original problem. We use rectangles and not simplices for the dissection of the range of the scenario process. While for a fixed stage, if the ξ_1 takes values in $\mathbb{R}^m$, then the range may be dissected into simplices, if we consider two subsequent stages, we always use rectangles, i.e., sets of the form $S_1 \times S_2$, where S_1 and S_2 are simplices. However, to make the notation simple, we consider here the case in which all ξ_t are univariate, i.e., the simplices are intervals. If $[a_1, a_2]$ is an interval for ξ_1 and $[b_1, b_2]$ is an interval for ξ_2, then $[a_1, a_2] \times [b_1, b_2]$ is the rectangle for $[\xi_1, \xi_2]$. Notice that points in the interior of rectangles do not have a unique representation as a convex combination of the corners, but for nonanticipative reasons we use product weights, as is explained later.

We review here a basic approximation result proved in [26], which will be needed later.

THEOREM 1.1. *Suppose that $Q(x_0, \xi_1, x_1, \ldots, \xi_T, x_T)$ is convex in all x and Lipschitz with Lipschitz constant L in ξ. Then the approximation error expressed in terms of the absolute difference in optimal values $|v(P) - v(\tilde{P})|$ can be bounded by*

$$(1.3) \qquad\qquad |v(P) - v(\tilde{P})| \leq L \cdot d(P, \tilde{P}),$$

where $d(P, \tilde{P})$ is the nested distance between the two scenario models P and $\tilde{P}$.

More generally, if the expectation $\mathbb{E}$ in (1.1) is replaced by a distortion functional $\mathcal{R}(Y) = \int_0^1 F_Y^{-1}(u)\, h(u)\, du$, with F_Y being the distribution function of Y and h being the nonnegative distortion function, then the assertion changes to

$$(1.4) \qquad\qquad |v(P) - v(\tilde{P})| \leq L \cdot \sup_{0 \leq u \leq 1} h(u) \cdot d(P, \tilde{P}).$$

Remark. The (upper) average value-at-risk (conditional value-at-risk, expected shortfall)

$$AV@R_\alpha(Y) = \min \left\{ q + \frac{1}{1-\alpha} \mathbb{E}[Y - q]_+ : q \in \mathbb{R} \right\}$$

is a distortion functional with

$$h(u) = \frac{1}{1-\alpha} \mathbb{1}_{\alpha \leq u \leq 1},$$

where $\mathbb{1}$ is the indicator function and thus in (1.4) we have $\sup_{0 \leq u \leq 1} h(u) = (1-\alpha)^{-1}$.

For the exact definition and properties of the nested distance $d(P, \tilde{P})$ we refer the reader to the book [26]. The nested distance is based on minimal transportation costs between the scenario processes ξ and $\tilde{\xi}$. Let us mention here that *any* feasible transportation plan between the two processes leads to an upper bound in (1.3). Thus a possible way of obtaining upper and lower bounds is as follows:
  1. Find a finite scenario process $\tilde{\xi}$ with distribution $\tilde{P}$ and a feasible[4] transportation plan π between the infinite process ξ and the finite process $\tilde{\xi}$.
  2. Solve the finite problem and find $v(\tilde{P})$.

---

[4]A transportation plan is feasible if it respects the filtration structure; see [26].

3. According to Theorem 1.1 the bounds are then given by

$$(1.5) \qquad v(\tilde{P}) - L \cdot d_\pi(P, \tilde{P}) \le v(P) \le v(\tilde{P}) + L \cdot d_\pi(P, \tilde{P}),$$

where d_π is the distance calculated with the transportation plan π. Obviously the best bounds are obtained for the optimal transportation plan, which defines the nested distance, but any other transportation plan also leads to guaranteed bounds.

However, it turns out that the bounds given in (1.5) are not very tight because they do not use the stochastic dominance properties of probability measures, when they are available. In this paper, however, we exploit these properties extensively.

The paper is organized as follows. In section 2 we describe the principles of bounding for two-stage stochastic optimization problems. Section 3 contains the main results for the multistage situation based on first-order and convex-order stochastic dominance. An example of the construction of upper and lower trees is contained in section 4. In section 5 we illustrate how to close the gap between upper and lower bounds. Finally, we report numerical results on a multistage production problem in section 6 and conclude the paper with section 7.

2. Bounding two-stage stochastic optimization problems. We consider two-stage stochastic optimization problems where the cost function in (1.1) is of the form $Q(x_0, \xi_1, x_1)$ containing just one single random variable $\xi \equiv \xi_1$. We assume here and in the following that the decisions x take values in $\mathbb{R}^d$, while the random variables ξ take values in $\mathbb{R}^m$.

We recall the definition of first-order stochastic dominance and convex dominance that will be useful for providing the bounds proposed in the paper.

DEFINITION 2.1 (stochastic dominance). *Let P, $\tilde{P}$, and $\bar{P}$ be probability distributions on $\mathbb{R}^m$. Consider the following stochastic dominance relations.*
  (i) First-order stochastic dominance (FSD). *P is dominated by $\tilde{P}$ in the first-order sense, and we write*

$$P \prec_{FSD} \tilde{P},$$

if $\int f(u)\, dP(u) \le \int f(u)\, d\tilde{P}(u)$ for all nondecreasing integrable real-valued functions f, i.e., for functions f for which $u' \le u''$ (componentwise) implies that $f(u') \le f(u'')$.
  (ii) Convex stochastic dominance (CXD). *P is dominated by $\bar{P}$ in the convex-order sense, and we write*

$$P \prec_{CXD} \bar{P},$$

if $\int f(u)\, dP(u) \le \int f(u)\, d\bar{P}(u)$ for all convex integrable f.

The relation $\prec_{CXD}$ is also known by the names *Bishop–de Leeuw ordering* or *Lorenz dominance*. More details about order relations can be found in [24]. Typically, probabilities being smaller in convex ordering can be obtained by concentrating the mass on the expectation (by virtue of using Jensen's inequality [14]) and probabilities being larger in convex ordering are obtained by moving the masses in a mean-preserving way to the boundaries of the convex hull of the support of P (the balayage as is done in the Edmundson–Madansky inequality [18, 19]). These simple (and well-known) facts are the basis of the results of this paper.

2.1. Bounds based on convex order for two-stage stochastic optimization problems.

LEMMA 2.2. *Suppose the probability measure P can be written as $P = \sum_i w_i P_i$, where w_i are nonnegative weights with $\sum_i w_i = 1$ and P_i are probability measures. Then*

$$(2.1) \qquad \underline{P} := \sum_i w_i \delta_{z_i} \prec_{CXD} P,$$

where δ_{z_i} is the point mass associated with the barycenter $z_i = \mathbb{E}(P_i)$ of P_i.

Proof. Let f be convex and integrable. Then

$$\int f(u)\, dP(u) = \sum_i w_i \int f(u) dP_i(u) \geq \sum_i w_i f(\mathbb{E}(P_i)) = \int f(u)\, d\underline{P}(u). \qquad \square$$

Thus, if the support of the probability P, say Ξ, is partitioned into a finite union of disjoint sets, $\Xi = \bigcup_i A_i$, and P_i are the conditional probabilities given A_i, i.e., $P_i(B) = P(B \cap A_i)/p_i$ with $p_i = P(A_i)$, and if $z_i = \mathbb{E}(P_i)$ are the barycenters, then $\sum_i p_i \delta_{z_i} \prec_{CXD} P$ (see Figure 2.1 for an example).
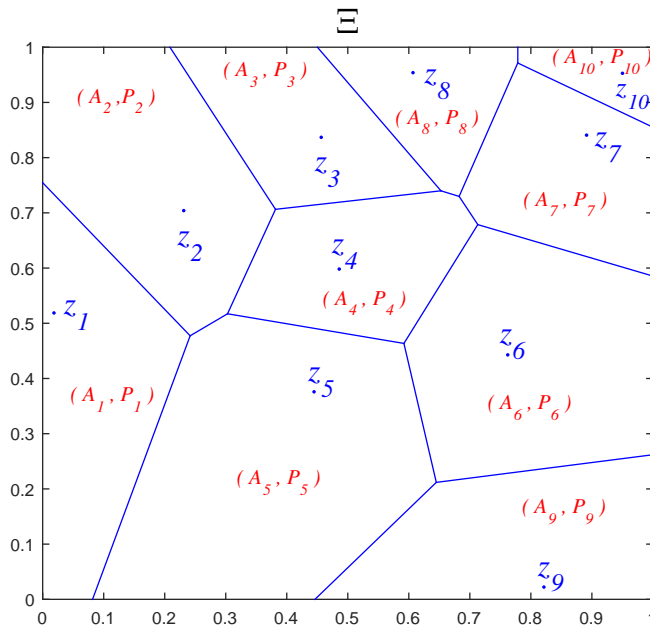

FIG. 2.1. *Example of a partition of a 2-dimensional support Ξ of P into a finite union of disjoint sets A_i such $\Xi = \bigcup_i A_i$ with expectation $\mathbb{E}(P_i)$.*

To get the inverse relation, suppose that the support of P is contained in the union of closed convex polyhedral sets A_i such that their interiors are disjoint. A_i is the convex hull of its extremal points, say $\{e_1^i, \ldots, e_{K_i}^i\}$. Each point u in the set A_i can be written as

$$(2.2) \qquad u = \sum_k w_k^i(u) e_k^i$$

with $w_k^i(u) \geq 0$ and $\sum_k w_k^i(u) = 1$. If the A_i's are simplices, then the representation in (2.2) is unique, but uniqueness is not required here (see Figure 2.2 for an example).
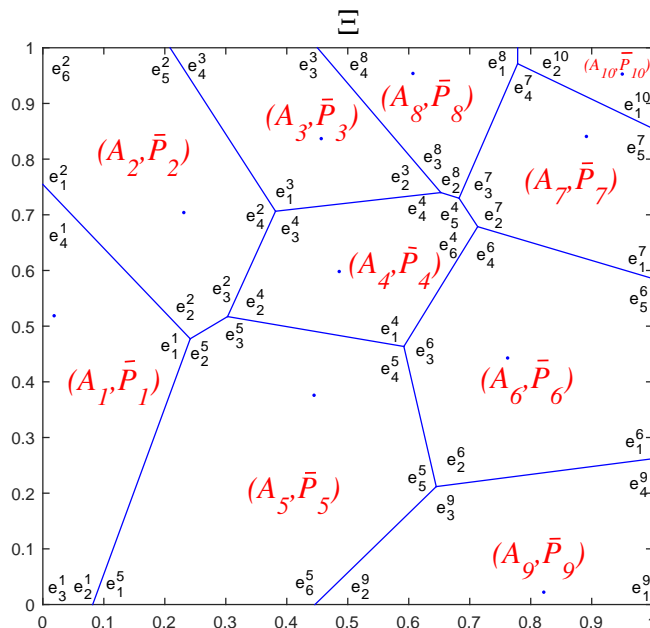

FIG. 2.2. *Example of a partition of a 2-dimensional support Ξ of P into a finite union of disjoint sets A_i such $\Xi = \bigcup_i A_i$ with extremal points $\{e_1^i, \ldots, e_{k_i}^i\}$ and $\bar{P}_i$ a discrete nonnegative measure sitting on the extremals of A_i.*

LEMMA 2.3. *Let $P_i^u = \sum_k w_k^i(u)\delta_{e_k^i}$[5] (where the weights are defined in (2.2)) and let $\bar{P}_i = \int_{A_i} P_i^u \, dP(u)$, which is a discrete nonnegative measure sitting on the extremals of A_i. Notice that the barycenter of P_i^u is $u \in A_i$. Let $\bar{P} = \sum_i \bar{P}_i$. Then $\bar{P}$ is a probability measure with the property that $P \prec_{CXD} \bar{P}$.*

Proof. Let f be convex and integrable. Then

$$\int f(u) \, d\bar{P}(u) = \sum_i \int_{A_i} f(u) \, d\bar{P}_i(u) = \sum_i \int_{A_i} \sum_k f(e_k^i) w_k^i(u) dP(u)$$

$$\geq \sum_i \int_{A_i} f\left(\sum_k w_k^i(u)e_k^i\right) dP(u) = \sum_i \int_{A_i} f(u) \, dP(u)$$

$$= \int f(u) dP(u). \qquad \square$$

Remark (some facts about convex order on intervals). Let the probability P be concentrated on the interval $[a_1, a_2]$ and have mean μ. It is well known that the one-point lower bound $\underline{P}$ and upper bound $\overline{P}$ are

$$(2.3) \qquad \underline{P} = \delta_\mu \qquad \text{and} \qquad \overline{P} = \frac{a_2 - \mu}{a_2 - a_1}\delta_{a_1} + \frac{\mu - a_1}{a_2 - a_1}\delta_{a_2}.$$

[5]As before δ_y is the point mass at y.

Notice that, for any affine function f,

$$\int f(u)\,d\underline{P}(u) = \int f(u)\,dP(u) = \int f(u)\,d\overline{P}(u).$$

One might conjecture that there may be nice lower approximations based on convex order such that equality holds even for quadratic functions, possibly by allowing the lower bound to sit on two points. The answer however is negative, as can be seen from the following example.

Example 1. Let P be the uniform distribution on $[0,1]$. In order that a discrete distribution in $[0,1]$ has second moment $1/3$ (as the uniform distribution) it has to sit with probability $1/2$ on the points $1/2 - \sqrt{1/12}$ and $1/2 + \sqrt{1/12}$. Call this distribution P_1. However, P_1 is not dominated according to the convex stochastic sense by P, since taking the function $f(u) = (1 - 2u)_+$ we have

$$\int_0^1 (1 - 2u)_+\,dP_1 = 0.288 > \int_0^1 (1 - 2u)_+\,dP(u) = 0.25.$$

One may also see that lower bounds for $\int f(u)\,dP(u)$, under additional constraints than just the expectation constraint, may depend on f. Consider the following example.

Example 2. Let the set of admissible probabilities P be discrete distributions sitting on at most three points in $[0,1]$ such that the expectation is $1/2$ and the second moment is $1/3$. Denote by $\#(\mathrm{supp}(P))$ the cardinality of the support of P. We may show that the problem

$$\mathrm{argmin}\left\{ \int_0^1 u^4\,dP(u) : \#(\mathrm{supp}(P)) \le 3, \int_0^1 u\,dP(u) = 1/2; \int_0^1 u^2\,dP(u) = 1/3 \right\}$$

has as solution $P = 1/2\delta_a + 1/2\delta_{1-a}$ with $a = 1/2 - \sqrt{1/12}$. On the other hand, the problem

$$\mathrm{argmin}\left\{ \int_0^1 -\sin(u)\,dP(u) : \right.$$
$$\left. \#(\mathrm{supp}(P)) \le 3, \int_0^1 u\,dP(u) = 1/2; \int_0^1 u^2\,dP(u) = 1/3 \right\}$$

has as a solution $P = 1/6\delta_0 + 2/3\delta_{1/2} + 1/6\delta_1$.

This means that solving the generalized moment problem as in [4, 6, 7] will give a bound for a specific fixed integrand, but not a dominating measure as in our approach.

Coming back to the case of a probability P on a general interval $[a_1, a_2]$, if in addition to the constraint that the expectation is μ one requires that the respective lower and upper probabilities $\underline{P}$ and $\overline{P}$ satisfy

$$\int_{a_1}^{a_2} |u - \mu|\,d\underline{P}(u) = \int_{a_1}^{a_2} |u - \mu|\,dP(u) = \int_{a_1}^{a_2} |u - \mu|\,d\overline{P} = d$$

and

$$\int_{\mu}^{a_2} d\underline{P}(u) = \int_{\mu}^{a_2} dP(u) = \int_{\mu}^{a_2} d\overline{P} = \beta,$$

then the solution is easily found: one has to consider the subintervals $[a_1, \mu]$ and $[\mu, a_2]$ and find the two lower and upper bounds as in (2.3) for the two intervals and put them together for a lower bound (which then sits on two points) and an upper bound (which then sits on three points). Doing so, the lower bound is

$$\underline{P} = (1 - \beta)\delta_{\mu - \frac{d}{2(1-\beta)}} + \beta\delta_{\mu + \frac{d}{2\beta}},$$

while the upper bound is

$$\overline{P} = \frac{d}{2(\mu - a_1)}\delta_{a_1} + \left(1 - \frac{d}{2(\mu - a_1)} - \frac{d}{2(a_2 - \mu)}\right)\delta_\mu + \frac{d}{2(a_2 - \mu)}\delta_{a_2}.$$

See Ben-Tal and Hochman [1] for the same result obtained in a different way. Notice that the refined bounds are simply obtained by dissecting the interval $[a_1, a_2]$ into two subintervals $[a_1, \mu]$ and $[\mu, a_2]$. Thus the refined bounds proposed in [1] appear in a natural way when the interval dissection is refined, and therefore they are included in our approach.

Remark (orderings of probability measures and risk functionals). If f is nondecreasing and real valued, then $P \prec_{FSD} \tilde{P}$ implies that $P^f \prec_{FSD} \tilde{P}^f$, where P^f and $\tilde{P}^f$ are the respective image measures under f. Thus, if a risk measure ρ is monotonic w.r.t. $\prec_{FSD}$, then $P \prec_{FSD} \tilde{P}$ and $u \mapsto Q(x, u)$ nondecreasing for all x implies that $\rho_P(Q(x, \cdot)) \le \rho_{\tilde{P}}(Q(x, \cdot))$. Hence also $\inf_x \rho_P(Q(x, \cdot)) \le \inf_x \rho_{\tilde{P}}(Q(x, \cdot))$.

Likewise, if f is convex and real valued, then $P \prec_{CXD} \bar{P}$ implies that $P^f \prec_{CXD} \bar{P}^f$, where P^f and $\bar{P}^f$ are the respective image measures under f. Thus, if a risk measure ρ is monotonic w.r.t. second-order stochastic dominance, then $P \prec_{CXD} P'$ and $u \mapsto Q(x, u)$ convex and nondecreasing for all x implies that $\rho_P(Q(x, \cdot)) \le \rho_{\bar{P}}(Q(x, \cdot))$. Hence, also $\inf_x \rho_P(Q(x, \cdot)) \le \inf_x \rho_{\bar{P}}(Q(x, \cdot))$.

3. Bounding multistage stochastic optimization problems. We consider now a multistage stochastic optimization problem of the form (1.1), where the uncertainty is described by a stochastic process $\xi = (\xi_1, \ldots, \xi_T)$. This process is characterized by P_1, the distribution of ξ_1, and the conditional distributions $\xi_t | (\xi_1 = u_1, \ldots, \xi_{t-1} = u_{t-1})$ for all $t > 1$, denoted by $P_t(\cdot | u_1, \ldots, u_{t-1})$. As already mentioned in the introduction, our goal is

(i) to find tree processes $\bar{P}$ and $\underline{P}$ such that with the value function v as in (1.1) we have

$$v(\underline{P}) \le v(P) \le v(\bar{P}),$$

or with a possible small correction term ϵ,

$$v(\underline{P}) - \epsilon \le v(P) \le v(\bar{P}) + \epsilon;$$

(ii) to be able to construct refinements of these bounds by considering bushier trees, if the gap is considered too large;

(iii) to find approximate solutions for the infinite problem (1.1) based in the solutions of the finite problem without solving the infinite problem.

3.1. Bounds based on first-order dominance. Stochastic first-order dominance given in Definition 2.1 may be broken down in a multistage setting to the conditional distributions. To this end, we introduce the following definition.

DEFINITION 3.1. *We say that a process ξ is totally monotone if the conditional distributions satisfy*

$$\xi_t|(\xi_1 = u'_1, \ldots, \xi_{t-1} = u'_{t-1}) \prec_{FSD} \xi_t|(\xi_1 = u''_1, \ldots, \xi_{t-1} = u''_{t-1})$$

whenever $u'_1 \leq u''_1, \ldots, u'_{t-1} \leq u''_{t-1}$.[6]

LEMMA 3.2. *Let the two processes ξ and $\tilde{\xi}$ be totally monotone. Let in addition*

$$(3.1) \qquad \xi_t|(\xi_1 = u_1, \ldots, \xi_{t-1} = u_{t-1}) \prec_{FSD} \tilde{\xi}_t|(\tilde{\xi}_1 = u_1, \ldots, \tilde{\xi}_{t-1} = u_{t-1}).$$

Then $P \prec_{FSD} \tilde{P}$.

Proof. Let f be monotonic in all arguments and let P and $\tilde{P}$ be the two probability distributions associated with the totally monotone processes ξ and $\tilde{\xi}$, which satisfy (3.1). Consider

$$f_{T-1}(u_1, \ldots, u_{T-1}) := \int f(u_1, \ldots, u_T) \, dP(u_T|u_1, \ldots, u_{T-1})$$

and

$$\tilde{f}_{T-1}(u_1, \ldots, u_{T-1}) := \int f(u_1, \ldots, u_T) \, d\tilde{P}(u_T|u_1, \ldots, u_{T-1}).$$

Then f_{T-1} and $\tilde{f}_{T-1}$ are monotonic in $(u_1, \ldots, u_T)$ and by assumption $f_{T-1} \leq \tilde{f}_{T-1}$. With a similar argument,

$$f_{T-2}(u_1, \ldots, u_{T-2}) := \int f(u_1, \ldots, u_{T-1}) \, dP(u_{T-1}|u_1, \ldots, u_{T-2}),$$

and the analogously defined $\tilde{f}_{T-2}$ are again monotonic and satisfy $f_{T-2} \leq \tilde{f}_{T-2}$. Continuing the integration to the end, one gets that

$$\int f(\cdot) \, dP(\cdot) \leq \int f(\cdot) \, d\tilde{P}(\cdot). \qquad \square$$

If the cost function $Q(x_0, \xi_1, x_1, \ldots, \xi_T, x_T)$ is monotonic in $(\xi_1, \ldots, \xi_T)$, we can apply Lemma (3.2) to problem (1.1) to construct upper bounds.

Example 3. Consider the following multistage stochastic optimization problem with linear constraints in x:

$$\min \left\{ c_0(x_0) + \mathbb{E}\left[\sum_{t=1}^{T} c_t(x_t, \xi_t) \right] : x \in \mathbb{X} \right\},$$

----

[6]These are stronger assumptions than those introduced in the previous papers [13] and [17].

where the feasible set $\mathbb{X}$ is given by

$$
\begin{aligned}
W_0 x_0 &\geq h_0, \\
A_1 x_0 + W_1 x_1 &\geq h_1(\xi_1), \\
A_2 x_1 + W_2 x_2 &\geq h_2(\xi_2), \\
&\ \ \vdots \\
A_T x_{T-1} + W_T x_T &\geq h_T(\xi_T), \\
x_1 &\lhd \mathcal{F}_1, \\
&\ \ \vdots \\
x_T &\lhd \mathcal{F}_T.
\end{aligned}
\tag{3.2}
$$

Then if $u \mapsto h_t(u)$ and $u \mapsto c_t(x_t, u)$ are monotonically nondecreasing, then the cost function $Q(x_0, \xi_1, x_1, \ldots, \xi_T, x_T)$ is monotonic in $(\xi_1, \ldots, \xi_T)$. In addition, if c_0 and c_t are monotonic in x and W_t are nonnegative, then $Q(x_0, \xi_1, x_1, \ldots, \xi_T, x_T)$ is also monotonic in x. For later use we may also state that if $(x, u) \mapsto c_t(x, u)$ and $u \mapsto h_t(u)$ are convex for all t, then the cost function $Q(x_0, \xi_1, x_1, \ldots, \xi_T, x_T)$ is jointly convex in all arguments.

In the following, we elaborate the method for a three stage model; the generalization to more stages can be done analogously, but the notation is more involved. Suppose that (ξ_1, ξ_2) take their values in a rectangle $\Xi = [L_1, U_1] \times [L_2, U_2]$. Let $L_1 = a_1 < a_2 < \cdots < a_{m_a+1} = U_1$ and, for each $i = 1, \ldots, m_a + 1$, let $L_2 = b_{i,1} < b_{i,2} < \cdots < b_{i,m_b+1} = U_2$. Define the rectangles $A_{i,j} = [a_i, a_{i+1}] \times [b_{i,j}, b_{i,j+1}]$, $i = 1, \ldots, m_a$, $j = 1, \ldots, m_b$. We assume that the scenario distribution (ξ_1, ξ_2) has a density and therefore it does not matter that the rectangles are not disjoint. Define $p_{i,j} = P(A_{i,j})$. Let the finite tree process $\tilde{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2)$ take in each rectangle $A_{i,j}$ the upper value $\tilde{\xi}_1 = a_{i+1}$ with probability $\sum_j p_{i,j}$ and the value on the right $\tilde{\xi}_2 = b_{i,j+1}$ with probability $p_{i,j} / \sum_j p_{i,j}$ conditional on $\tilde{\xi}_1 = a_{i+1}$. Then $\tilde{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2) := (a_{i+1}, b_{i,j+1})$ is a finite tree process $\tilde{P}$ for which we solve the basic problem (1.1). Let the respective solutions be $\tilde{x}_1^i$ and $\tilde{x}_2^{i,j}$. We extend this to a decision function on Ξ by setting

$$
\tilde{x}_1(u_1) = \tilde{x}_1^i \qquad \text{when} \qquad u_1 \in A_i := [a_i, a_{i+1}]
$$

and

$$
\tilde{x}_2(u_1, u_2) = \tilde{x}_2^{i,j} \qquad \text{when} \qquad (u_1, u_2) \in A_{i,j}.
$$

We get

$$
\begin{aligned}
v(\tilde{P}) &= \int Q(x_0, u_1, \tilde{x}_1, u_2, \tilde{x}_2)\, d\tilde{P}(u_1, u_2) \\
&= \sum_{i,j} p_{i,j} Q(x_0, a_{i+1}, \tilde{x}_1^i, b_{i,j+1}, \tilde{x}_2^{i,j}) \\
&\geq \sum_{i,j} \int_{A_{i,j}} Q(x_0, u_1, \tilde{x}_1^i, u_2, \tilde{x}_2^{i,j})\, dP(u_1, u_2) \\
&= \sum_{i,j} \int_{A_{i,j}} Q(x_0, u_1, \tilde{x}_1(u_1), u_2, \tilde{x}_2(u_1, u_2))\, dP(u_1, u_2) \\
&= \int (Q(x_0, u_1, \tilde{x}_1(u_1), u_2, \tilde{x}_2(u_1, u_2))\, dP(u_1, u_2) \geq v(P).
\end{aligned}
$$

For establishing a lower bound we use the same setup as before. Let the finite tree process $\underset{\sim}{\xi} = (\underset{\sim}{\xi}_1, \underset{\sim}{\xi}_2)$ take in each rectangle $A_{i,j}$ the lower value $\underset{\sim}{\xi}_1 = a_i$ with probability $\sum_j p_{i,j}$ and the value on the left $\underset{\sim}{\xi}_2 = b_{i,j}$ with probability $p_{i,j}/\sum_j p_{i,j}$ conditional on $\underset{\sim}{\xi}_1 = a_i$. Then $\underset{\sim}{\xi} = (\underset{\sim}{\xi}_1, \underset{\sim}{\xi}_2) := (a_i, b_{i,j})$ is a finite tree process $\underset{\sim}{P}$ for which we solve the basic problem (1.1). Furthermore, we have to assume that the function $Q(x_0, \xi_1, x_2, \ldots, \xi_T, x_T)$ is also monotonic in the x-variables. If $x_1^*(u_1)$ and $x_2^*(u_1, u_2)$ are the solutions of the infinite problem, we set

$$x_1^+(a_i) = \min_{u_1 \in A_i} x_1^*(u_1),$$
$$x_2^+(a_i, b_{i,j}) = \min_{(u_1, u_2) \in A_{i,j}} x_2^*(u_1, u_2).$$

Since these are feasible but not necessarily optimal solutions, we get, by the assumed monotonicity,

$$v(P) = \int Q(x_0, u_1, x_1^*(u_1), u_2, x_2^*(u_1, u_2)) \, dP(u_1, u_2)$$
$$\geq \sum_{i,j} \int_{A_{i,j}} Q(x_0, u_1, x_1^+(a_i), u_2, x_2^+(a_i, b_{i,j})) \, dP(u_1, u_2) \geq v(\underset{\sim}{P}).$$

Notice that we have not only bounded the objective value, but also found an ϵ-solution for the continuous problem if ϵ is the gap between the upper and the lower bound.

3.2. Bounds based on convex dominance. In this section we provide bounds based on convex dominance for multistage stochastic programs. One might conjecture that for two scenario processes, $P_t(u_t|u_1, \ldots, u_t) \prec_{CXD} \bar{P}_t(u_t|u_1, \ldots, u_t)$ for all t and all $(u_1, \ldots, u_T)$ is sufficient to entail $P \prec_{CXD} \bar{P}$. However this is not true, as the following example shows.

Example 4. Let $\xi_1 \sim N(0, \sigma_1^2)$ and $\xi_2|\xi_1 \sim N(\exp(-\xi_1^2/4), \sigma_1^2)$. Similarly, $\bar{\xi}_1 \sim N(0, \sigma_2^2)$ and $\bar{\xi}_2|\bar{\xi}_1 \sim N(\exp(-\bar{\xi}_1^2/4), \sigma_2^2)$. If $\sigma_1 < \sigma_2$, then $\xi_1 \prec_{CXD} \bar{\xi}_1$ and also $(\xi_2|\xi_1 = x) \prec_{CXD} (\bar{\xi}_2|\bar{\xi}_1 = x)$ for all x. But $(\xi_1, \xi_2) \nprec_{CXD} (\bar{\xi}_1, \bar{\xi}_2)$, as can be seen from the second moments of ξ_2. Choose, e.g., $\sigma_1 = 1/2$, $\sigma_2 = 1$. Then

$$\mathbb{E}(\xi_2^2) = \sigma_1^2 + \sqrt{1/\sigma_1^2 + 1/\sigma_1^4} = 4.7221 > 2.4142 = \sigma_2^2 + \sqrt{1/\sigma_2^2 + 1/\sigma_2^4} = \mathbb{E}(\bar{\xi}_2^2).$$

Notice also that a convex-order dominating discrete probability cannot be found by choosing dominating discretizations for first components and for all conditional distributions of the second component and concatenating them together, as as the next example shows.

Example 5. Suppose that ξ_1 is distributed according to Uniform$[0, 1]$ and that $\xi_2|\xi_1$ is distributed according to Uniform$[\xi_1(1 - \xi_1), \xi_1(1 - \xi_1) + 1]$. Let $\bar{\xi}_1$ take the values 0 and 1, each with probability $1/2$. Then $\bar{\xi}_1$ dominates ξ_1 in convex order. Likewise, for each u, let $\bar{\xi}_2(u)$ take the values $u(1 - u)$ and $u(1 - u) + 1$, each with probability $1/2$. Then the conditional distributions $\xi_2|\xi_1 = u$ are dominated by $\bar{\xi}_2(u)$ for all u. But, if concatenating $\bar{\xi}_1$ with the conditional distributions $\bar{\xi}_2(u)$, only the conditional distributions for $u = 0$ and $u = 1$ are used, and one obtains that $(\bar{\xi}_1, \bar{\xi}_2)$ has a distribution that sits on all 4 edges of the unit square with equal probabilities $1/4$. But this is not a convex dominant of (ξ_1, ξ_2), since

$$\mathbb{E}(\xi_2^2) = 16/30 > \mathbb{E}(\bar{\xi}_2^2) = 1/2.$$

The same example shows that concatenating lower bound approximations for the conditional probabilities does not lead to a lower bound approximation for the total probability. That is why, in our procedure, the full multistage sample is used to generate the bounding trees and not just the conditional distributions. Notice that this does not provide a counterexample for [11, 12], since in our example ξ_2 depends on ξ_1 in a very nonlinear way. On the contrary, in [11] the authors assume that the ξ_t are mutually independent, and the authors of [12] assume that the ξ_t follow an autoregressive process (linear dependence on the past).

Again, we concentrate here on three-stage programs, noticing that the generalization to more stages is straightforward, but needs complicated notation. Let $(\xi_1, \xi_2) \mapsto Q(x_0, \xi_1, x_1, \xi_2, x_2)$ be a convex function and let P be a probability measure on a bounded rectangle in $\mathbb{R}^2$. Notice that the extension to $\mathbb{R}^m \times \mathbb{R}^m$ is obvious, so we omit it. As in (1.1), $\mathcal{F}_1$ is the σ-algebra generated by the first component in $\mathbb{R}^2$ and $\mathcal{F}$ is a Borel σ-algebra of $\mathbb{R}^2$. Our problem is to find

$$(3.3) \qquad v(P) = \min_{x_0, x_1, x_2} \mathbb{E}_P[Q(x_0, \xi_1, x_1(\xi_1), \xi_2, x_2(\xi_1, \xi_2))],$$

where x_0 is deterministic, x_1 is measurable w.r.t. $\mathcal{F}_1$, and x_2 is measurable w.r.t. $\mathcal{F}$.

3.2.1. Upper tree approximation based on convex-order stochastic dominance. Upper bounds for minimization problems are always easy to obtain, since every feasible solution constitutes an upper bound. If the basic problem P contains continuous distributions, but the approximating problem $\bar{P}$ is discrete, then one has to construct a feasible solution for P out of one for $\bar{P}$. Suppose that $\bar{P}$ is a scenario tree with values z_1^i in the second stage and $z_2^{i,j}$ in the third stage and $\bar{x}_1^i$ and $\bar{x}_2^{i,j}$ are its discrete solutions of the problem with $\bar{P}$ as the distribution of the scenario process. Then by any reasonable extension function one may construct a solution for the continuous problem (3.3), for instance by setting

$$\bar{x}_1(\xi_1) = \bar{x}_1^i \qquad \text{if } z_1^i \text{ is the point that is closest to } \xi_1,$$
$$\bar{x}_2(\xi_1, \xi_2) = \bar{x}_2^{i,j} \qquad \text{if } (z_1^i, z_2^{i,j}) \text{ is the point that is closest to } (\xi_1, \xi_2).$$

Obviously,

$$\min_{x_0, x_1(\cdot), x_2(\cdot)} \mathbb{E}_P[Q(x_0, \xi_1, x_1(\xi_1), \xi_2, x_2(\xi_1, \xi_2))] \le \mathbb{E}_{\bar{P}}[Q(x_0, \xi_1, \bar{x}_1(\xi_1), \xi_2, \bar{x}_2(\xi_1, \xi_2))].$$

That is, any extension of a solution of any tree process $\bar{P}$ leads to a valid upper bound. However, notice that one has to evaluate the objective function for the scenario process P and the solution $\bar{x}$ in order to get the value of the upper bound.

We aim, however, at finding an upper bound, which can be calculated on a finite tree without evaluating the continuous problem. A construction similar to the one for $\bar{P}$ in the two-stage case may be used.

Suppose that (ξ_1, ξ_2) take their values in a rectangle $\Xi = [L_1, U_1] \times [L_2, U_2]$. Let $L_1 = a_1 < a_2 < \cdots < a_{m_a+1} = U_1$ and, for each i, let $L_2 = b_{i,1} < b_{i,2} < \cdots < b_{i,m_b+1} = U_2$ as before. Define the rectangles $A_{i,j} = [a_i, a_{i+1}] \times [b_{i,j}, b_{i,j+1}]$, $i = 1, \ldots, m_a$, $j = 1, \ldots, m_b$. We assume that the scenario distribution (ξ_1, ξ_2) has a density and therefore it does not matter that the rectangles are not disjoint. From now on, we use the notation $e_k^{i,j}$ for the four corners $k = 1, \ldots, 4$ of the rectangle $A_{i,j}$, irrespective whether these are a_i's or $b_{i,j}$'s.

Notice there are infinitely many ways to represent a point in a rectangle as a convex representation of the corners. However, if the rectangle represents the intervals

for the second and the third stages, we will use a product form for the weights, which implies that the weights for ξ_1 do not depend on ξ_2: let a typical rectangle be $A_{i,j} = [a_i, a_{i+1}] \times [b_{i,j}, b_{i,j+1}]$. Then, defining as in (2.3)

$$w_{a_i}(\xi_1) = \frac{a_{i+1} - \xi_1}{a_{i+1} - a_i}, \qquad w_{a_{i+1}}(\xi_1) = \frac{\xi_1 - a_i}{a_{i+1} - a_i},$$

for $a_i \leq \xi_1 \leq a_{i+1}$ one has $\xi_1 = w_{a_i}(\xi_1)a_i + w_{a_{i+1}}(\xi_1)a_{i+1}$, and similarly for $b_{i,j} \leq \xi_2 \leq b_{i,j+1}$. Then, for (ξ_1, ξ_2) in the rectangle $A_{i,j}$ the weights are

$$\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = w_{a_i}(\xi_1)w_{b_{i,j}}(\xi_2) \begin{pmatrix} a_i \\ b_{i,j} \end{pmatrix} + w_{a_{i+1}}(\xi_1)w_{b_{i,j}}(\xi_2) \begin{pmatrix} a_{i+1} \\ b_{i,j} \end{pmatrix}$$

$$(3.4) \qquad + w_{a_i}(\xi_1)w_{b_{i,j+1}}(\xi_2) \begin{pmatrix} a_i \\ b_{i,j+1} \end{pmatrix} + w_{a_{i+1}}(\xi_1)w_{b_{i,j+1}}(\xi_2) \begin{pmatrix} a_{i+1} \\ b_{i,j+1} \end{pmatrix}.$$

In general notation, denoting the four corners by $e_1^{i,j}$, $e_2^{i,j}$, $e_3^{i,j}$, $e_4^{i,j}$ we may equivalently write

$$(\xi_1, \xi_2) = \sum_{i=1}^{4} w_{e_k^{i,j}}(\xi_1, \xi_2)\delta_{e_k^{i,j}}.$$

The contribution of the probability P with the weights on the corners is

$$w_k^{i,j} = \int_{A_{i,j}} w_{e_k^{i,j}}(u_1, u_2) \, dP(u_1, u_2).$$

The numerical calculation of these integrals is nearly impossible, and therefore we replace the exact calculation by a Monte Carlo integration: if $(\xi^{(i)}, \ldots, \xi^{(N)})$ are N replications of samples from the stochastic process ξ, then the weights are defined as

$$w_k^{i,j} = \sum_{n=1}^{N} w_{e_k^{i,j}}(\xi_1^{(n)}, \xi_2^{(n)})\mathbb{1}_{(\xi_1^{(n)}, \xi_2^{(n)}) \in A_{i,j}}.$$

Let $P_{i,j}$ be the distribution P conditioned on the set $A_{i,j}$, i.e.,

$$P_{i,j}(D) = \frac{1}{p_{i,j}}P(A_{i,j} \cap D),$$

with $p_{i,j} = P(A_{i,j})$ and D any measurable set. We have that

$$P = \sum_{i,j} p_{i,j}P_{i,j}.$$

Let $\bar{P}_{i,j}$ be a probability measure sitting on the extremals (the corners) of $A_{i,j}$ such that $P_{i,j} \prec_{CXD} \bar{P}_{i,j}$, as is constructed in Lemma 2.3. Then $P \prec_{CXD} \bar{P}$ by the following lemma.

LEMMA 3.3. *If $P_{i,j} \prec_{CXD} \bar{P}_{i,j}$ for all i, j and*

$$P = \sum_{i,j} p_{i,j}P_{i,j} \qquad and \qquad \bar{P} = \sum_{i,j} p_{i,j}\bar{P}_{i,j},$$

then

$$P \prec_{CXD} \bar{P}.$$

Proof. Let f be convex. Then

$$\mathbb{E}_P(f) = \sum_{i,j} p_{i,j} \mathbb{E}_{P_{i,j}}(f) \leq \sum_{i,j} p_{i,j} \mathbb{E}_{\bar{P}_{i,j}}(f) = \mathbb{E}_{\bar{P}}(f). \qquad \square$$

$\bar{P}$ is a tree process and one may find the solution of the optimization problem on this tree. Let $x_1^*(e_k^i)$ and $x_2^*(e_k^{i,j})$ be a solution of this problem sitting on the nodes e_k^i, $e_k^{i,j}$ of the tree. We construct a continuous extension of this solution, i.e., a nonanticipative solution $x_1^*(\xi_1)$ and $x_2^*(\xi_1, \xi_2)$ for arbitrary values of the process (ξ_1, ξ_2). If $(\xi_1, \xi_2) \in A_{i,j}$, then $(\xi_1, \xi_2) = \sum_k w_k^{i,j}(\xi_1, \xi_2) e_k^{i,j}$, while by construction $\xi_1 = \sum_k w_k^i(\xi_1) e_k^i$ with $w_k^i(\xi_1) = \sum_j w_k^{i,j}(\xi_1, \xi_2)$. Set $\bar{x}_1(\xi_1) = \sum_k w_k^i(\xi_1) x_1^*(e_k^i)$ and $\bar{x}_2(\xi_1, \xi_2) = \sum_k w_k^{i,j}(\xi_1, \xi_2) x_2^*(e_k^{i,j})$. By construction

$$v(\bar{P}) = \mathbb{E}_{\bar{P}}[Q(x_0, \xi_1, x_1^*, \xi_2, x_2^*)] \geq \mathbb{E}_P[Q(x_0, \xi_1, \bar{x}_1(\xi_1), \xi_2, \bar{x}_2(\xi_1, \xi_2))].$$

That is, the constructed continuous solution leads to a lower bound for the objective function. By replacing this solution with the optimal solution $x_1^*(\xi_1)$, $x_2^*(\xi_1, \xi_2)$ of the continuous problem, we see that $v(\bar{P}) \geq v(P)$, showing that we found a valid upper bound. Notice that the constructed continuous solution is an ϵ-solution for the continuous problem, if ϵ is the gap between the upper and the lower bound.

We finally illustrate the fact that upper bounds may be wrongly calculated when one concentrates the tree construction on conditional distributions by the following simple, but instructive, example.

Example 6. Let ξ_1 be distributed according to Uniform$[0,1]$ and let the conditional distribution $\xi_2|\xi_1$ be distributed according to Beta$(0.01 + 0.36 \cdot \xi_1(1 - \xi_1), 0.3)$. We consider the optimization problem

$$(3.5) \qquad v(P) := \min_x \{\mathbb{E}(\xi_2 - x)^2 : x \text{ is measurable w.r.t. } \xi_1\}.$$

Evidently, the minimal value $v(P)$ is the expected conditional variance of ξ_2 given ξ_1, i.e., $\mathbb{E}(Var(\xi_2|\xi_1))$. A numerical calculation shows that $v(P) = 0.1065$. Now, finding the upper tree $\tilde{P}$, given by $\tilde{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2)$ by finding first the upper probability $\tilde{\xi}_1$ sitting on 0 or 1 and then the upper probability $\tilde{\xi}_2$ given that $\tilde{\xi}_1 = 0$ or 1, results in

$$P(\tilde{\xi}_1 = 0) = P(\tilde{\xi}_1 = 1) = 0.5, \quad P(\tilde{\xi}_2 = 0|\tilde{\xi}_1) = 0.9678, \quad P(\tilde{\xi}_2 = 1|\tilde{\xi}_1) = 0.0322.$$

However, constructing the tree $\bar{P}$, given by $\bar{\xi} = (\bar{\xi}_1, \bar{\xi}_2)$ according to our method, we get

$$P(\bar{\xi}_1 = 0) = P(\bar{\xi}_1 = 1) = 0.5, \quad P(\bar{\xi}_2 = 0|\bar{\xi}_1) = 0.5655, \quad P(\bar{\xi}_2 = 1|\bar{\xi}_1) = 0.4345.$$

Solving problem (3.5) for the tree $\tilde{P}$ gives the value $v(\tilde{P}) = 0.0312$, whereas solving the same problem for the tree $\bar{P}$ leads to the value $v(\bar{P}) = 0.2437$. Thus, the wrong construction $\tilde{P}$ does not result in an upper bound, while the correct construction $\bar{P}$ does.

Again notice that Example 6 does not provide a counterexample to [11, 12] since in [11] the authors assume that the ξ_t are mutually independent, and the authors of [12] assume that the ξ_t follow an autoregressive process (linear dependence on the past).

3.2.2. Lower tree approximation based on convex-order stochastic dominance. The problem of finding lower bounds is slightly more involved than finding the upper bounds. We refer to the same construction of rectangles as before. Let $z^{i,j} = (z_1^{i,j}, z_2^{i,j})$ be the barycenter of $P_{i,j}$ and let $\underline{P}_{i,j} = \delta_{z^{i,j}}$.

By construction, $\underline{P}_{i,j} \prec_{CXD} P_{i,j}$ and by Lemma 3.3, $\underline{P} \prec_{CXD} P$.

Let $\mathcal{F}^A$ be the σ-algebra generated by the sets $A_{i,j}$. Notice that for all integrable functions f,

$$\mathbb{E}[f(\xi_1, \xi_2)|\mathcal{F}^A] = \sum_{i,j} \mathbb{E}_{P_{i,j}}(f) \mathbb{1}_{A_{i,j}}.$$

We now distinguish two cases.
- *Case* I. Conditionally on A_i, ξ_2 is independent of ξ_1.
- *Case* NI. The independence assumption is not satisfied.

Case I. In this case, the conditional expectations given $A_{i,j}$ of all functions that only depend on the first component ξ_1 do not depend on j; in particular, $z_1^{i,j} = z_1^i$ do not depend on j.

The following proposition holds true in this case.

PROPOSITION 3.4. *Let $\underline{P}$ be the tree constructed using z_1^i and $z_2^{i,j}$ with scenario probabilities $p_{i,j}$. Then $v(\underline{P}) \leq v(P)$.*

Proof. Let $x_1(\xi_1)$, $x_2(\xi_1, \xi_2)$ be a solution of the infinite problem. Then

$$
\begin{aligned}
v(P) &= \mathbb{E}_P \left[Q(x_0, \xi_1, x_1(\xi_1), \xi_2, x_2(\xi_1, \xi_2)) \right] \\
&= \mathbb{E}_P \left[\mathbb{E}_P \left[Q\left(x_0, \xi_1, x_1(\xi_1), \xi_2, x_2(\xi_1, \xi_2)\right) | \mathcal{F}^A \right] \right] \\
&\geq \mathbb{E}_P \left[Q\left(x_0, \mathbb{E}(\xi_1|\mathcal{F}^A), \mathbb{E}(x_1|\mathcal{F}^A), \mathbb{E}(\xi_2|\mathcal{F}^A), \mathbb{E}(x_2|\mathcal{F}^A)\right) \right] \\
&= \mathbb{E}_{\underline{P}} \left[Q\left(x_0, z_1^i, x_1^i, z_2^{i,j}, x_2^{i,j}\right) \right] \\
&\geq v(\underline{P}),
\end{aligned}
$$

where $x_1^i = \mathbb{E}(\xi_1|A_i)$ and $x_2^{i,j} = \mathbb{E}(\xi_2|A_{i,j})$. The last inequality comes from the fact that we have constructed a feasible solution for $\underline{P}$, but this need not be the optimal one. However, its optimal solution leads to a further lower bound for the problem with P as the scenario process. Therefore $\underline{P}$ is a lower tree approximation. □

Case NI. If the independence assumption is not valid, we use a different approach.

Assumption A. Since $z \mapsto Q(x, z)$ is convex and finitely valued, for every point $\bar{z}$ we have

$$Q(x, z) \geq Q(x, \bar{z}) + \langle q(\bar{z}|x), z - \bar{z} \rangle,$$

where $q(\bar{z}|x)$ is a subgradient of $z \mapsto Q(x, z)$ at $\bar{z}$ for fixed x. We assume that there is a uniform bound $C(\bar{z})$ for the norm of the subgradients $q(\bar{z}|x)$, i.e.,

$$\sup_x \|q(\bar{z}|x)\| \leq C(\bar{z}),$$

such that $C(\xi_1)$ is integrable.

Under Assumption A, for all $\bar{\xi}$ we have

$$
\begin{aligned}
\mathbb{E}[Q(x, \xi)] &\geq \mathbb{E}[Q(x, \bar{\xi})] + \mathbb{E}\langle q(\bar{\xi}|x), (\xi - \bar{\xi}) \rangle \\
&\geq \mathbb{E}[Q(x, \bar{\xi})] - \mathbb{E}(C(\bar{\xi})) \cdot \sup_\xi \|\bar{\xi} - \xi\|.
\end{aligned}
$$
(3.6)

The case of constraints of the form $(x, z) \in B$ needs a little more attention. Suppose that $Q(x, z) = Q_0(x, z) + \mathbb{I}_B(x, z)$, where $Q_0(x, z)$ is convex and finitely valued and B is a convex set.

If $(x, \bar{z}) \in B$ and $q(x, \bar{z})$ is a subgradient of $z \mapsto Q(x, z)$ at $\bar{z}$ for fixed x, then $Q(x, z) \geq Q(x, \bar{z}) + \langle q(\bar{z}|x), z - \bar{z} \rangle$ even if $(x, z) \notin B$. Therefore (3.6) is a valid lower bound for all values of ξ.

A similar argument holds for the three-stage case: if $z_1 \mapsto Q(x_0, z_1, x_1, z_2, x_2)$ is convex, with subgradient $q(z_1|x_0, x_1, x_2, z_2)$, then

$$
\begin{aligned}
&\mathbb{E}_P[Q(x_0, \xi_1, x_1, \xi_2, x_2)] \\
&\geq \mathbb{E}[Q(x_0, \bar{\xi}_1, x_1, \xi_2, x_2)] + \mathbb{E}[\langle q(\bar{\xi}_1|x_0, x_1, x_2, \xi_2), (\xi_1 - \bar{\xi}_1) \rangle] \\
&\geq \mathbb{E}[Q(x_0, \bar{\xi}_1, x_1, \xi_2, x_2)] - \mathbb{E}_P(C(\bar{\xi}_1)) \cdot \sup_{\xi_1} \|\xi_1 - \bar{\xi}_1\|,
\end{aligned}
$$

where $\bar{\xi}_1 = \mathbb{E}(\xi_1 | \mathcal{F}^A)$.

Notice that instead of the inequality

$$
(3.7) \qquad \mathbb{E}[\langle q(\bar{\xi}_1|x_0, x_1, x_2, \xi_2), (\xi_1 - \bar{\xi}_1) \rangle] \leq \mathbb{E}_P(C(\bar{\xi}_1)) \cdot \sup_{\xi_1} \|\xi_1 - \bar{\xi}_1\|,
$$

one could also use the inequality

$$
(3.8) \qquad \mathbb{E}[\langle q(\bar{\xi}_1|x_0, x_1, x_2, \xi_2), (\xi_1 - \bar{\xi}_1) \rangle] \leq [\mathbb{E}_P(C^p(\bar{\xi}_1))]^{1/p} \cdot [\mathbb{E}_P(\|\xi_1 - \bar{\xi}_1\|^r)]^{1/r}
$$

for $1/p + 1/r = 1$.

For the construction of the lower approximating tree, as before let $z^{i,j} = (z_1^{i,j}, z_2^{i,j})$ be the barycenters of $P_{i,j}$. However, since the lower approximation has to be a tree, we set

$$
\bar{z}_1^i = \sum_{j=1}^{m_b} p_{i,j} z_2^{i,j}.
$$

Let $\underline{P}$ be the tree constructed using $\bar{z}_1^i$ as first stage values and and $z_2^{i,j}$ as second stage values with scenario probabilities $p_{i,j}$. The notation of this tree as well as of the decision tree is shown in Figure 3.1.

PROPOSITION 3.5. *Let $\underline{P}$ be the tree constructed using $\bar{z}_1^i$ as first stage values and and $z_2^{i,j}$ as second stage values with scenario probabilities $p_{i,j}$. For the tree process $\underline{P}$ we have that*

$$
v(\underline{P}) - \sum_i p_i \, C(\bar{z}_1^i) \cdot \sup \|\bar{z}_1^i - z_1^{i,j}\| \leq v(P).
$$

Proof. As in Proposition 3.4, we have that

$$
\begin{aligned}
&\mathbb{E}_P[Q(x_0, \xi_1, x_1(\xi_1), \xi_2, x_2(\xi_1, \xi_2))] \\
&\quad \geq \mathbb{E}_{\underline{P}}[Q(x_0, z_1^{i,j}, x_1^i, z_2^{i,j}, x_2^{i,j})] \\
&\quad = \mathbb{E}_{\underline{P}}[Q(x_0, \bar{z}_1^i, x_1^i, z_2^{i,j}, x_2^{i,j})] + \mathbb{E}_{\underline{P}}[Q(x_0, z_1^{i,j}, x_1^i, z_2^{i,j}, x_2^{i,j})] \\
&\qquad - \mathbb{E}_{\underline{P}}[Q(x_0, \bar{z}_1^i, x_1^i, z_2^{i,j}, x_2^{i,j})] \\
&\quad \geq \mathbb{E}_{\underline{P}}[Q(x_0, \bar{z}_1^i, x_1^i, z_2^{i,j}, x_2^{i,j})] - \sum_i p_i \, C(\bar{z}_1^i) \cdot \sup \|\bar{z}_1^i - z_1^{i,j}\|.
\end{aligned}
$$

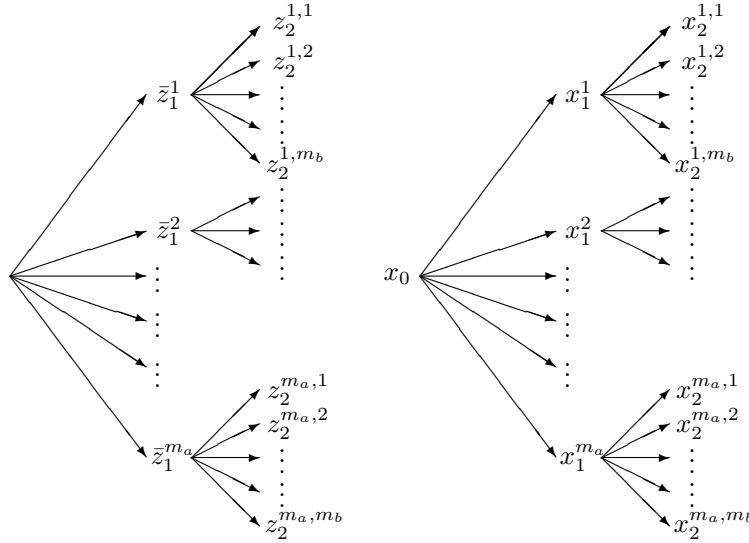Here we have used (3.7), but alternatively one may also use (3.8).

FIG. 3.1. *Left: the notation of the constructed scenario tree. Right: the pertaining decision tree.*

If the $z_1^{i,j}$ do not coincide for different j but fixed i, then the correction term

$$\sum_i p_i \, C(\bar{z}_1^i) \cdot \sup \|\bar{z}_1^i - z_1^{i,j}\|,$$

has to be subtracted. Otherwise the correction term disappears. Notice that the correction term is small if the diameters of the rectangles are small. □

4. Lower and upper scenario trees construction: An example. In order to demonstrate the approach proposed in section 3 with a simple example, assume that the distribution P of the scenario process is given by (ξ_1, ξ_2), where $\xi_1 \sim P_1$ is distributed according to a Beta$(2,2)$ distribution and $\xi_2|\xi_1 \sim P_2(\cdot|\xi_1)$ is conditionally given ξ_1 distributed according to Beta$(2, \frac{1.4-0.8\cdot\xi_1}{0.3+0.4\cdot\xi_1})$. A sample of 5000 points distributed according to P is shown in Figure 4.1.

For given integers m_a and m_b we define the sets $A_{i,j}$ as the rectangles with vertices $[(\frac{i-1}{m_a}, \frac{j-1}{m_b}), (\frac{i}{m_a}, \frac{j-1}{m_b}), (\frac{i-1}{m_a}, \frac{j}{m_b}), (\frac{i}{m_a}, \frac{j}{m_b})]$ for $i = 1, \ldots, m_a+1$, $j = 1, \ldots, m_b+1$. The upper approximation sits on the $(m_a + 1) \cdot (m_b + 1)$ points $(i/m_a, j/m_b)$. The lower approximation sits on some barycenters of the $A_{i,j}$. The probabilities are $p_{i,j} = P(A_{i,j})$.

If (u_1, u_2) is a point in the rectangle $A_{i,j}$ defined before, then let $P_{i,j}(u_1, u_2)$ be a probability measure sitting on the vertices with probability

$$p\left(\frac{i-1}{m_a}, \frac{j-1}{m_b}\right) = \frac{\frac{i}{m_a} - u_1}{\frac{i}{m_a} - \frac{i-1}{m_a}} \cdot \frac{\frac{j}{m_b} - u_2}{\frac{j}{m_b} - \frac{j-1}{m_b}} = (i - m_a \cdot u_1)(j - m_b \cdot u_2),$$

$$p\left(\frac{i-1}{m_a}, \frac{j}{m_b}\right) = \frac{\frac{i}{m_a} - u_1}{\frac{i}{m_a} - \frac{i-1}{m_a}} \cdot \frac{u_2 - \frac{j-1}{m_b}}{\frac{j}{m_b} - \frac{j-1}{m_b}} = (i - m_a \cdot u_1)(m_b \cdot u_2 - j + 1),$$
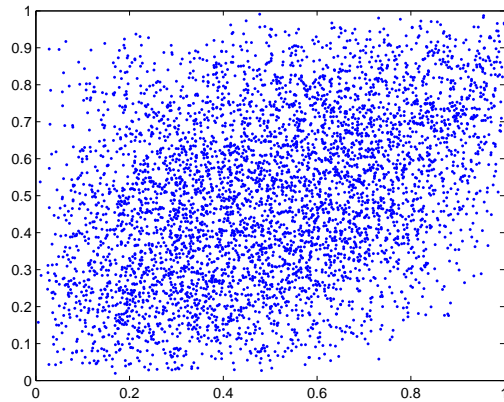
FIG. 4.1. 5000 *points distributed according to the distribution* P *of the scenario process* (ξ_1, ξ_2), *with* $\xi_1 \sim P_1$ *distributed according to a* Beta$(2,2)$ *distribution and* $\xi_2 | \xi_1 \sim P_2(\cdot | \xi_1)$ *according to a* Beta$(2, \frac{1.4 - 0.8 \cdot \xi_1}{0.3 + 0.4 \cdot \xi_1})$ *distribution.*
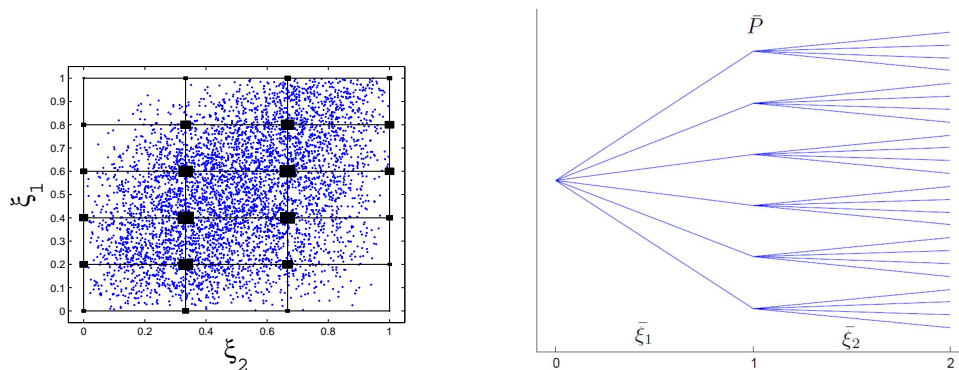

FIG. 4.2. *The upper approximation* $\bar{P}$ *based on convex stochastic dominance (left) and the corresponding scenario tree structure (right).*

$$p\left(\frac{i}{m_a}, \frac{j-1}{m_b}\right) = \frac{u_1 - \frac{i-1}{m_a}}{\frac{i}{m_a} - \frac{i-1}{m_a}} \cdot \frac{\frac{j}{m_b} - u_2}{\frac{j}{m_b} - \frac{j-1}{m_b}} = (m_a \cdot u_1 - i + 1)(j - m_b \cdot u_2),$$

$$p\left(\frac{i}{m_a}, \frac{j}{m_b}\right) = \frac{u_1 - \frac{i-1}{m_a}}{\frac{i}{m_a} - \frac{i-1}{m_a}} \cdot \frac{u_2 - \frac{j-1}{m_b}}{\frac{j}{m_b} - \frac{j-1}{m_b}} = (m_a \cdot u_1 - i + 1)(m_b \cdot u_2 - j + 1).$$

Notice that the expectation of $P_{i,j}(u_1, u_2)$ is (u_1, u_2).

In order to estimate the upper and lower approximations $\bar{P}$ and $\underline{P}$, we use a large sample of N random deviates $(\xi_1^{(n)}, \xi_2^{(n)})$. Set

$$\bar{P} = \frac{1}{N} \sum_{n=1}^{N} P_{i,j}(\xi_1^{(n)}, \xi_2^{(n)}) \cdot \mathbb{1}_{(\xi_1^{(n)}, \xi_2^{(n)}) \in A_{i,j}}.$$

$\bar{P}$ is a finite process, and defines a tree process, which is the upper approximation. Figure 4.2 shows a construction of an upper approximation $\bar{P}$ based on convex stochastic dominance with the corresponding scenario tree structure with $m_a = 5$ and $m_b = 3$.
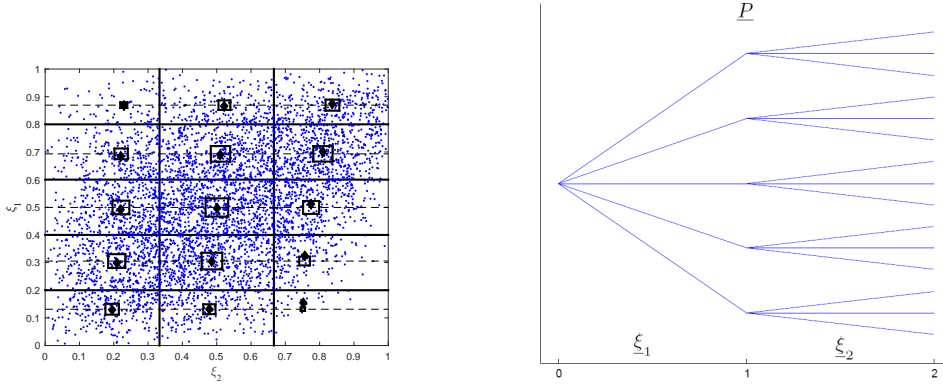
FIG. 4.3. *The barycenters (black diamonds) and the modified barycenters (black squares) of the lower approximation tree based on convex stochastic dominance (left) and the corresponding scenario tree structure (right).*

For the lower bound, the generation algorithm is a little more complicated. We sample $(\xi_1^{(n)}, \xi_2^{(n)})$, $n = 1, \ldots, N$, from P and set

$$n_{i,j} = \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{(\xi_1^{(n)}, \xi_2^{(n)}) \in A_{i,j}},$$

$$p_{i,j} = n_{i,j}/N,$$

$$z_1^{i,j} = \frac{1}{n_{i,j}} \sum_{n=1}^{N} \xi_1^{(n)} \mathbb{1}_{(\xi_1^{(n)}, \xi_2^{(n)}) \in A_{i,j}},$$

$$z_2^{i,j} = \frac{1}{n_{i,j}} \sum_{n=1}^{N} \xi_2^{(n)} \mathbb{1}_{(\xi_1^{(n)}, \xi_2^{(n)}) \in A_{i,j}},$$

$$\bar{z}_1^i = \frac{\sum_j p_{i,j} z_1^{i,j}}{\sum_j p_{i,j}}.$$

Then $\underline{P}$ is defined as

$$\underline{P} = \sum_{i,j} p_{i,j} \delta_{(\bar{z}_1^i, z_2^{i,j})}.$$

$\underline{P}$ can be represented as a tree. Arcs (i, j) for which $p_{i,j} = 0$ can be eliminated. Figure 4.3 shows the barycenters $(z_1^{i,j}, z_2^{i,j})$ (black diamonds) and the modified barycenters $(\bar{z}_1^i, z_2^{i,j})$ (black squares) for the choice $m_a = 5$, $m_b = 3$, and the distribution as in Figure 4.1.

Finally Figure 4.4 shows an upper approximation $\tilde{P}$ based on first-order stochastic dominance with the corresponding scenario tree structure as described in section 3.1. Similarly, a lower approximation $\underline{P}$ based on first-order stochastic dominance can be obtained by putting the weights to the left-lower corner of each rectangles in which the support has been dissected.

5. Closing the gap between upper and lower bounds. By refinement of the discretization, i.e., by letting the diameter of the boxes go to zero, we may get an arbitrarily small gap between the upper and lower bounds under the additional assumption of a growth condition. This is illustrated by the following proposition.
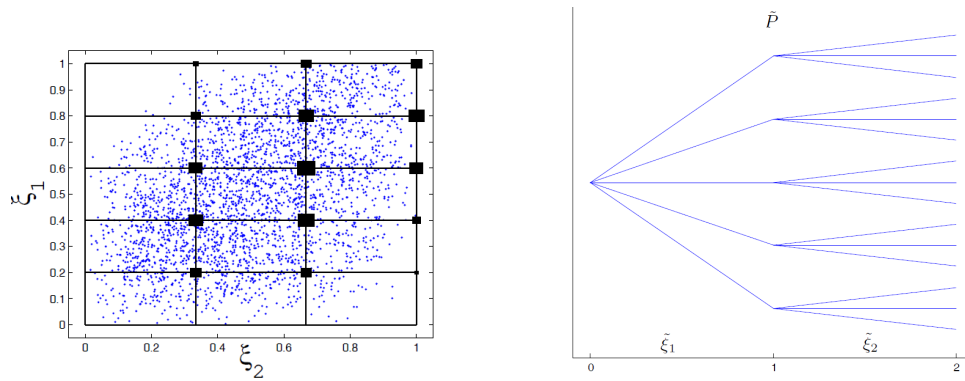
FIG. 4.4. *The upper approximation $\tilde{P}$ based on first-order stochastic dominance (left) and the corresponding scenario tree structure (right).*

PROPOSITION 5.1. *Suppose that the convex cost function Q satisfies the growth conditions*

$$Q(x,z) \leq Q(x,\bar{z}) + \langle q(\bar{z}|x), z - \bar{z} \rangle + c\|z - \bar{z}\|^{\gamma}$$

for $\gamma > 0$ uniformly for all $x \in \mathbb{X}$, where $q(\bar{z}|x)$ is a subgradient of $z \mapsto Q(x,z)$ at $\bar{z}$. Let A be a closed convex polyhedron and let z_A be its barycenter under P. If $\bar{P}$ is any probability distribution whose support is contained in A and which also has barycenter z_A, then

$$0 \leq \int_A Q(x,z)\,d\bar{P}(z) - Q(x,z_A) \leq c \cdot \sup_{z \in A} \|z - z_A\|^{\gamma},$$

uniformly for all $x \in \mathbb{X}$. Thus the difference between the upper and the lower bound can be controlled by making the diameter of A small.

Proof. By

$$Q(x,z) \leq Q(x,z_A) + \langle q(z_A|x), z - z_A \rangle + c \cdot \|z - z_A\|^{\gamma},$$

taking the integrals w.r.t. $\bar{P}$ one gets

$$0 \leq \int_A Q(x,z)\,d\bar{P}(z) - Q(x,z_A) \leq c \cdot \sup_{z \in A} \|z - z_A\|^{\gamma}.$$

Thus, by making the diameter of all boxes smaller than $\epsilon/c^{1/\gamma}$, the gap between the upper and the lower approximation will be smaller than ϵ, independently of the decisions x. Therefore the gap for the optimal values $v(\overline{P}) - v(\underline{P})$ is also smaller than ϵ. □

6. Case study: A multistage production problem. This section presents a simple multistage production problem adopted to test the bounds introduced before. The problem can be summarized as follows: consider a single product inventory system, which is comprised of a warehouse and a factory. The planning horizon is T periods. Random demands have to be satisfied from an inventory (the only random quantities in the model). If the random demand exceeds the stock, it will be satisfied by rapid orders from a different source, which come at a higher price. At each time step (stage), orders can be placed. The goal is to minimize the total production cost of the factory in the entire planning period.

Let us assume the following notation.

Deterministic parameters:

c_t is the cost of producing a unit of the product at the factory at time $t = 0, \ldots,$ $T - 1$;

b_t is the procurement cost from another retailer for a unit of product at time $t = 1, \ldots, T$;

s_t is the selling price at time $t = 1, \ldots, T$;

h_t is the inventory holding cost for positive inventory from time t to $t + 1$, $t = 0, \ldots, T - 1$;

d is the final value of the inventory;

$\mathcal{P}_t$ is the maximal production capacity of the factory at time $t = 0, \ldots, T - 1$;

v_0 is the amount of the product in the warehouse at the beginning of the period 1.

Stochastic scenario process:

ξ_t is the demand for the product at time $t = 1, \ldots, T$ and all the demand must be satisfied (the random scenario process);

ξ^t is the history of the demand for the product until time t.

Stochastic decision variables:

$x_t \geq 0$ is the amount of the product to be produced by the factory and used to satisfy the demand at time $t = 0, \ldots, T - 1$.

Auxiliary variables:

v_t the amount of the product in the warehouse after sales are effectuated at $t = 1, \ldots, T$.

Notice that if v_t is positive, $v_t = [v_t]_+$, an inventory holding cost $h_t \cdot [v_t]_+$ will be paid to carry the stock to the next step. If v_t is negative, $v_t = [v_t]_-$, a procurement cost $b_t [v_t]_-$ to buy extra stock from another retailer will be paid. The final stock is valuated with the value $d [v_T]_+$.

The problem can be modelled as follows:

$$\min \mathbb{E}\left[c_0 \cdot x_0 + h_0 \cdot v_0 + \sum_{t=1}^{T-1} c_t \cdot x_t(\xi^t) + \sum_{t=1}^{T-1} h_t \cdot [v_t(\xi^t)]_+ + \sum_{t=1}^{T} b_t \cdot [v_t(\xi^t)]_- \right.$$

$$(6.1) \qquad \left. - \sum_{t=1}^{T} s_t \cdot \xi_t - d \cdot [v_T(\xi^T)]_+ \right],$$

$$(6.2) \qquad \text{s.t. } 0 \leq x_0 \leq \mathcal{P}_0,$$

$$(6.3) \qquad 0 \leq x_t(\xi^t) \leq \mathcal{P}_t, \qquad t = 1, \ldots, T - 1,$$

$$(6.4) \qquad v_1(\xi_1) = v_0 + x_0 - \xi_1,$$

$$(6.5) \qquad v_{t+1}(\xi^{t+1}) = [v_t(\xi^t)]_+ + x_t(\xi^t) - \xi_{t+1}, \qquad t = 1, \ldots, T - 1,$$

$$(6.6) \qquad v_t(\xi^t) = [v_t(\xi^t)]_+ - [v_t(\xi^t)]_-, \qquad t = 1, \ldots, T,$$

$$(6.7) \qquad [v_t(\xi^t)]_+ \geq 0, \qquad t = 1, \ldots, T,$$

$$(6.8) \qquad [v_t(\xi^t)]_- \geq 0, \qquad t = 1, \ldots, T.$$

The objective function (6.1) denotes the expected total cost obtained from production, procurement from external retailers, and inventory holding while the last two terms are

the respective profits from selling and for the final value of the inventory. Constraints (6.2) and (6.3) impose lower and upper levels on the factory production. Finally, constraints (6.4), (6.5), (6.6), (6.7), and (6.8) define the dynamics of the inventory level and its definition.

6.1. Risk-aversion strategy: Including the average value-at-risk. Given the confidence level α, we now introduce into model (6.1)–(6.8) the (upper) average value-at-risk:

$$AV@R_\alpha = \min\left\{ y + \frac{1}{(1-\alpha)}\mathbb{E}([c_0 \cdot x_0 + h_0 \cdot v_0 + \sum_{t=1}^{T-1} c_t \cdot x_t(\xi^t) \right.$$
$$+ \sum_{t=1}^{T-1} h_t \cdot [v_t(\xi^t)]_+ + \sum_{t=1}^{T} b_t \cdot [v_t(\xi^t)]_- - \sum_{t=1}^{T} s_t \cdot \xi_t$$

(6.9)
$$\left. -d \cdot [v_T(\xi^T)]_+ - y]_+) : y \in \mathbb{R} \right\},$$

where y represents the value-at-risk ($V@R$). If $\alpha = 0$, then $AV@R_0$ equals the expectation and if $\alpha = 1$, then $AV@R_1$ is consistently defined as the essential supremum.

Introducing the auxiliary variable $u(\xi^T)$, the model (6.1)–(6.8) in a risk-aversion strategy becomes

(6.10)
$$\min y + \frac{1}{(1-\alpha)}\mathbb{E}u(\xi^T),$$

$$\text{s.t.} \quad u(\xi^T) \geq c_0 \cdot x_0 + h_0 \cdot v_0 + \sum_{t=1}^{T-1} c_t \cdot x_t(\xi^t)$$
$$+ \sum_{t=1}^{T-1} h_t \cdot [v_t(\xi^t)]_+ + \sum_{t=1}^{T} b_t \cdot [v_t(\xi^t)]_- - \sum_{t=1}^{T} s_t \cdot \xi_t$$

(6.11)
$$-d \cdot [v_T(\xi^T)]_+ - y,$$

(6.12)
$$0 \leq x_0 \leq \mathcal{P}_0,$$

(6.13)
$$0 \leq x_t(\xi^t) \leq \mathcal{P}_t, \qquad t = 1, \ldots, T-1,$$

(6.14)
$$v_1(\xi_1) = v_0 + x_0 - \xi_1,$$

(6.15)
$$v_{t+1}(\xi^{t+1}) = \left[v_t(\xi^t)\right]_+ + x_t(\xi^t) - \xi_{t+1}, \qquad t = 1, \ldots, T-1,$$

(6.16)
$$v_t(\xi^t) = \left[v_t(\xi^t)\right]_+ - \left[v_t(\xi^t)\right]_-, \qquad t = 1, \ldots, T,$$

(6.17)
$$\left[v_t(\xi^t)\right]_+ \geq 0, \qquad t = 1, \ldots, T,$$

(6.18)
$$\left[v_t(\xi^t)\right]_- \geq 0, \qquad t = 1, \ldots, T,$$

(6.19)
$$u(\xi^T) \geq 0.$$

6.2. Computation of bounds for a multistage risk-averse production problem. This section presents some computational tests on the three-stage ($T = 2$) risk-averse production problem. We assume that the distribution P of the demand scenario process is given by $\xi^2 = (\xi_1, \xi_2)$, where $\xi_1 \sim P_1$ is distributed according to a Beta(2, 2) distribution and $\xi_2|\xi_1 \sim P_2(\cdot|\xi_1)$ is conditionally given ξ_1 distributed

TABLE 6.1

Production price c_t, selling price s_t, inventory holding cost h_t from time t to time $t + 1$, and procurement cost b_t for extra stock from another retailer at time t.

t	c_t	s_t	h_t	b_t
0	3.5	–	2	–
1	3.6	10.7	1.9	8
2	–	10.5	–	8.1

TABLE 6.2

Scenario tree structures based on first-order stochastic dominance (FSD) and convex-order dominance (CXD) adopted to compute the bounds.

	Tree	Number of scenarios	Number of nodes
FSD-CXD	$T_{5,5}$	25	31
FSD-CXD	$T_{10,10}$	100	111
FSD-CXD	$T_{20,20}$	400	421
FSD	$T_{40,40}$	1600	1641
FSD	$T_{80,80}$	6400	6481
FSD	$T_{160,160}$	25600	25761

according to $\text{Beta}(2, (1.4 - 0.8 \cdot \xi_1)/(0.3 + 0.4 \cdot \xi_1))$ in the range $[0, 100]$. The maximal production capacity of the factory at each period $t = 0, 1$ is $\mathcal{P}_t = 567$ units. The initial inventory is $v_0 = 10$, the final value of the inventory is $d = 2$ per unit, and the values of production price c_t, selling price s_t, inventory holding cost h_t, and procurement cost b_t at time period t are presented in Table 6.1.

The linear problems derived from our case study were solved in the *AMPL* environment using the *MOSEK* solver with an interior-point algorithm. All the computations were performed on a 64-bit machine with 12 GB of RAM and a 2.90 GHz processor.

In order to find guaranteed bounds, we first consider two finite three-stage trees $\tilde{P} = (\tilde{\xi}_1, \tilde{\xi}_2)$ and $\underset{\sim}{P} = (\underset{\sim}{\xi}_1, \underset{\sim}{\xi}_2)$ having the same structure $T_{5,5}$: five branches from the root and five from each of the second-stage nodes resulting in $k = 5 \times 5 = 25$ scenarios and 31 nodes. The two finite scenario trees have been built according to first-order stochastic dominance, as described in section 3.1, providing respective upper and lower bounds: they are obtained by dissecting the support into 25 rectangles $A_{i,j}$, $i = 1, \ldots, 5$, $j = 1, \ldots, 5$, and putting the weights respectively to the left and lower corner (a_i, b_j) and to the up and right corner (a_{i+1}, b_{j+1}). Similarly, other pairs of finite scenario trees with bushier tree structures $T_{10,10}$, $T_{20,20}$, $T_{40,40}$, $T_{80,80}$, and $T_{160,160}$ have been considered (see Table 6.2 for details). Lower and upper bounds to the total cost of problem (6.10)–(6.19) by using the finite scenario trees based on first-order stochastic dominance (FSD) are reported in Tables 6.3, 6.4, and 6.5 and Figures 6.1 and 6.2. As expected, the worst lower and upper bounds are given by $T_{5,5}$ with an absolute gap $v(\tilde{P}) - v(\underset{\sim}{P})$ of 137.5 but requiring the lowest CPU time (0.0625 CPU seconds over 30 runs). Increasing the size of the scenario tree significantly improves the bounds, monotonically reaching lower values of gaps of up to 4.11 for the biggest scenario tree considered $T_{160,160}$ (see Figure 6.1 in the case of $\alpha = 0$, where the bounds are plotted for increasing values of complexity of calculation measured in CPU seconds). Similar results are obtained for different values

TABLE 6.3

Lower bound objective function values and complexity of calculation (in CPS seconds) of finite scenario tree structures, based on first-order stochastic dominance (FSD) for increasing values of α.

Trees	$v(\underline{P})$						CPU (s)
	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$	
$T_{5,5}^{FSD}$	361.66	379.61	414.36	450.22	479.13	514.65	0.062
$T_{10,10}^{FSD}$	399.09	415.49	447.27	479.49	514.06	565.42	0.078
$T_{20,20}^{FSD}$	416.66	433.89	464.34	495.87	531.56	582.08	0.093
$T_{40,40}^{FSD}$	424.98	441.91	472.85	503.94	539.45	589.18	0.156
$T_{80,80}^{FSD}$	427.86	445.24	477.25	509.58	545.95	593.42	0.375
$T_{160,160}^{FSD}$	430.71	448.02	479.51	511.47	547.34	596.37	3.406

TABLE 6.4

Upper bound objective function values and complexity of calculation (in CPS seconds) of finite scenario tree structures based on first-order stochastic dominance (FSD) for increasing values of α.

Trees	$v(\tilde{P})$						CPU (s)
	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$	
$T_{5,5}^{FSD}$	498.11	517.68	550.63	589.31	616.53	652.39	0.062
$T_{10,10}^{FSD}$	466.95	484.41	517.45	550.13	585.02	636.19	0.078
$T_{20,20}^{FSD}$	450.76	468.31	498.78	530.31	566.00	616.52	0.093
$T_{40,40}^{FSD}$	440.40	457.90	490.28	522.79	558.68	606.50	0.156
$T_{80,80}^{FSD}$	437.88	455.18	486.54	518.54	554.38	604.34	0.375
$T_{160,160}^{FSD}$	434.42	451.76	483.42	515.61	551.73	601.18	3.406

TABLE 6.5

Gaps of finite scenario tree structures based on first-order stochastic dominance (FSD) for increasing values of α.

Trees	$\frac{v(\tilde{P})-v(\underline{P})}{v(\underline{P})}$					
	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$
$T_{5,5}^{FSD}$	0.377	0.363	0.328	0.308	0.286	0.267
$T_{10,10}^{FSD}$	0.170	0.165	0.156	0.147	0.138	0.125
$T_{20,20}^{FSD}$	0.081	0.079	0.074	0.069	0.064	0.059
$T_{40,40}^{FSD}$	0.036	0.036	0.036	0.037	0.035	0.029
$T_{80,80}^{FSD}$	0.023	0.022	0.019	0.017	0.015	0.018
$T_{160,160}^{FSD}$	0.008	0.008	0.008	0.008	0.008	0.008

of confidence level α (see Figure 6.2). The time required to solve the problem (see last column of Tables 6.3 and 6.4) monotonically increases with the dimension of the tree, reaching the highest value for $T_{160,160}$ (3.40625 CPU seconds over 30 runs). Finally, average relative gaps $(v(\tilde{P}) - v(\underline{P}))/v(\underline{P})$ are reported in Table 6.5: as expected they improve monotonically with the number of scenarios in the trees, ranging from 32% for $T_{5,5}$ to 0.8% for $T_{160,160}$.
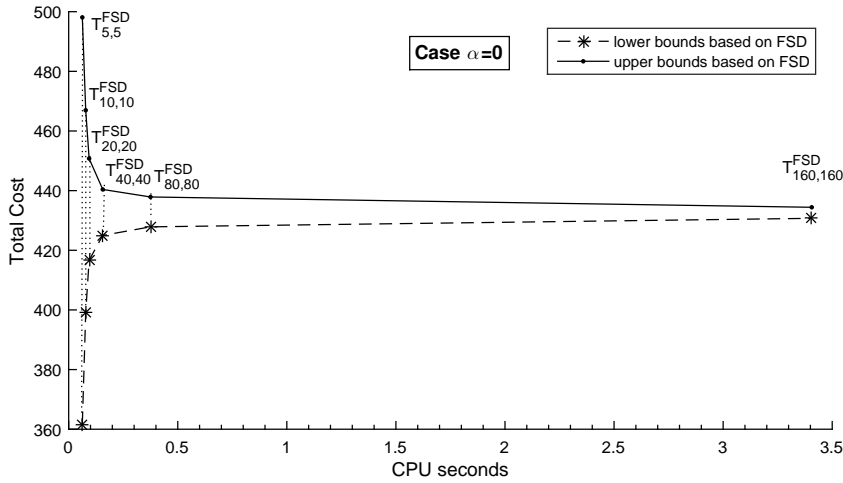
FIG. 6.1. *Lower and upper bounds on the total cost of problem* (6.10)–(6.19) *with confidence level* $\alpha = 0$ *obtained by using the finite scenario trees based on first-order stochastic dominance for increasing values of complexity of calculation measured in CPU seconds.*
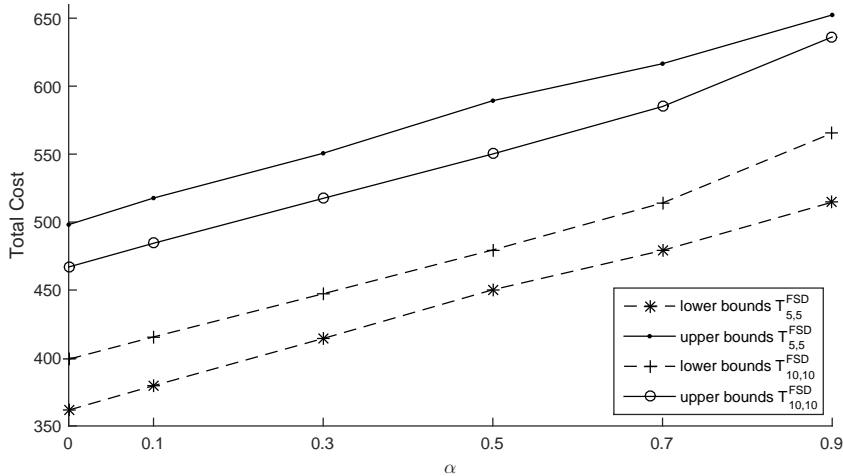

FIG. 6.2. *Lower and upper bounds on the total cost of problem* (6.10)–(6.19) *obtained by using the finite scenario trees* $T_{5,5}$ *and* $T_{10,10}$ *based on first-order stochastic dominance.*

We now consider lower and upper bounds built on convex stochastic dominance, as described in section 3.2. They are obtained by dissecting the support into $m_a \times m_b$ rectangles $A_{i,j}$, $i = 1, \ldots, m_a$, $j = 1, \ldots, m_b$, and respectively putting the weights to the barycenter and to the four corners. In this way the bounds can be calculated on two finite trees without evaluating the continuous problem.

Lower and upper bounds based on convex stochastic dominance (CXD) are reported in Tables 6.6, 6.7, and 6.8 and Figures 6.3 and 6.4. Since ξ_2 depends on ξ_1,

TABLE 6.6
Lower bound objective function values, errors $c_1 \sum_{i,j} p_{i,j} \cdot |\bar{z}_1^i - z_1^{i,j}|$, and complexity of calculation (in CPS seconds) of finite scenario tree structures based on convex stochastic dominance (CXD) for increasing values of α.

Trees	$v(\underline{P})$						Error	CPU (s)
	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$		
$\mathrm{T}_{5,5}^{CXD}$	424.74	442.55	473.21	509.15	535.18	571.15	0.0044	0.046
$\mathrm{T}_{10,10}^{CXD}$	431.40	448.12	479.49	511.74	543.62	592.49	0.0031	0.059
$\mathrm{T}_{20,20}^{CXD}$	431.40	448.21	479.81	512.43	549.41	597.81	0.0027	0.076

TABLE 6.7
Upper bound objective function values and complexity of calculation (in CPS seconds) of finite scenario tree structures based on convex stochastic dominance (CXD) for increasing values of α.

Trees	$v(\bar{P})$						CPU (s)
	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$	
$\mathrm{T}_{5,5}^{CXD}$	436.25	454.48	484.83	520.67	575.95	621.62	0.042
$\mathrm{T}_{10,10}^{CXD}$	431.64	448.50	479.82	513.74	553.09	604.95	0.054
$\mathrm{T}_{20,20}^{CXD}$	431.40	448.21	479.81	512.43	549.41	597.81	0.078

TABLE 6.8
Gaps of finite scenario tree structures based on convex dominance (CXD) for increasing values of α.

Trees	$\frac{v(\bar{P}) - v(\underline{P})}{v(\underline{P})}$					
	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$
$\mathrm{T}_{5,5}^{CXD}$	0.027	0.026	0.024	0.022	0.076	0.088
$\mathrm{T}_{10,10}^{CXD}$	0.0005	0.0008	0.0006	0.003	0.017	0.020
$\mathrm{T}_{20,20}^{CXD}$	0	0	0	0	0	0

the correction term described in section 3.2.2 for lower tree approximation should be computed (see the third column of Table 6.6). This is obtained as follows: problem (6.1)–(6.8) can be rewritten as

$$
\min \mathbb{E}\Bigg[c_0 \cdot x_0 + h_0 \cdot v_0 + \sum_{t=1}^{T-1} c_t \cdot x_t(\xi^t) + \sum_{t=1}^{T-1} h_t \cdot [v_t(\xi^t)]_+ + \sum_{t=1}^{T} b_t \cdot [v_t(\xi^t)]_-
$$

$$
- \sum_{t=1}^{T} s_t \cdot \xi_t - d \cdot [v_T(\xi^T)]_+ + \Psi[x_0, \ldots, x_T, \xi^T] \Bigg],
$$

where

$$
\Psi[x_0, \ldots, x_T, \xi^T] = \begin{cases} 0 & \text{if } (x_0, \ldots, x_T) \in \mathbb{X}, \\ \infty & \text{otherwise,} \end{cases}
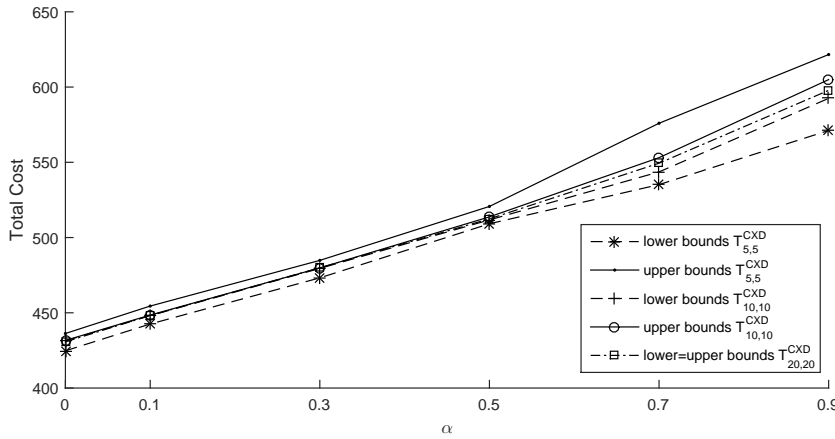$$

FIG. 6.3. *Lower and upper bounds to the total cost of problem* (6.10)–(6.19) *obtained by using the finite scenario trees* $T_{5,5}$, $T_{10,10}$, *and* $T_{20,20}$ *based on convex stochastic dominance (CXD).*

with

$$(6.20) \qquad \mathbb{X} := \begin{cases} 0 \le x_0 \le \mathcal{P}_0, \\ 0 \le x_t(\xi^t) \le \mathcal{P}_t, \qquad t = 1, \dots, T-1, \\ v_1(\xi_1) = v_0 + x_0 - \xi_1, \\ v_{t+1}(\xi^{t+1}) = \left[v_t(\xi^t)\right]_+ + x_t(\xi^t) - \xi_{t+1}, \qquad t = 1, \dots, T-1, \\ v_t(\xi^t) = \left[v_t(\xi^t)\right]_+ - \left[v_t(\xi^t)\right]_-, \qquad t = 1, \dots, T, \\ \left[v_t(\xi^t)\right]_+ \ge 0, \qquad t = 1, \dots, T, \\ \left[v_t(\xi^t)\right]_- \ge 0, \qquad t = 1, \dots, T. \end{cases}$$

Let $\underline{P}$ be the tree constructed using $\tilde{z}_1^i$ as first stage values and and $z_2^{i,j}$ as second stage values with scenario probabilities $p_{i,j}$. According to Proposition 3.5, we have that the error made by the tree process $\underline{P}$ for our three-stage production problem by collapsing $z_1^{i,j}$ in $\bar{z}_1^i$, $i = 1, \dots, m_a$, is $c_1 \sum_{i,j} p_{i,j} \cdot |\bar{z}_1^i - z_1^{i,j}|$. In the risk-averse case we need just to divide the previous expression by the confidence level $(1 - \alpha)$. Notice that if the demand at time 2 is independent of the demand at time 1, then the error is null. The absolute gap between CXD lower and upper bounds based on the simplest tree structure considered $T_{5,5}$ reduces considerably compared to the one obtained by first-order construction, passing, in the case of $\alpha = 0$, from 136.44 to 11.5 units. Increasing the size of the scenario tree to $T_{20,20}$ significantly improves the bounds closing the gap (see Figure 6.4) from 34.1 for FSD and taking approximately the same CPU time. Different values of confidence level α are considered in Figure 6.3 and relative gaps $(v(\bar{P}) - v(\underline{P}))/v(\underline{P})$ are reported in Table 6.8: results show that the average gap reduces considerably, passing from 4% with $T_{5,5}^{CXD}$ to 0% with $T_{20,20}^{CXD}$.

7. Conclusions. This paper develops lower and upper bounds for multistage stochastic programs based on first-order stochastic dominance and convex-order dominance of probability measures. The proposed method allows one to construct solutions for the infinite problem by considering finite tree approximations as proxies,
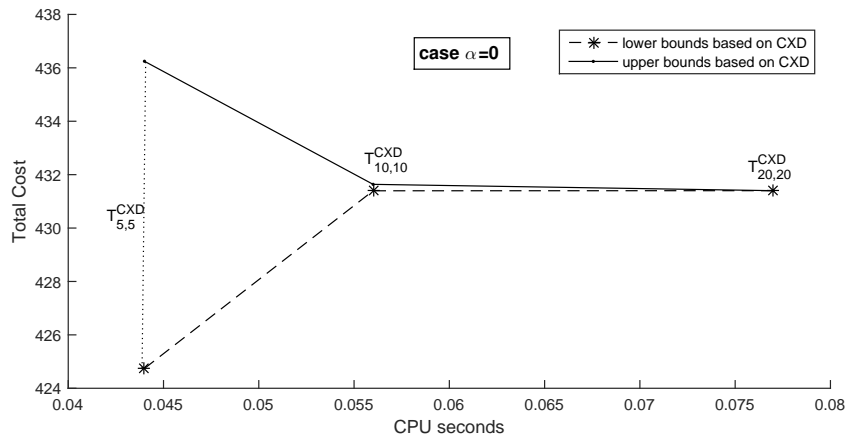
FIG. 6.4. *Lower and upper bounds to the total cost of problem* (6.10)–(6.19) *with* $\alpha = 0$ *obtained by using finite scenario trees based on convex dominance (CXD).*

and can even be made arbitrarily close by making the approximating trees bushier. For illustration, numerical results on a multistage risk-averse production problem are presented. Results show that the solutions based on convex-order dominance construction outperform the ones obtained by first-order stochastic dominance, closing the gap between upper and lower bounds within a limited computational complexity and simple scenario tree structures.

REFERENCES

[1] A. BEN-TAL AND E. HOCHMAN, *More bounds on the expectation of a convex function of a random variable*, J. Appl. Probab., 9 (1972), pp. 803–812.

[2] J. R. BIRGE, *The value of the stochastic solution in stochastic linear programs with fixed recourse*, Math. Program., 24 (1982), pp. 314–325.

[3] N. BOLAND, I. BAKIR, B. DANDURAND, AND A. ERERA, *Scenario Set Partition Dual Bounds for Multistage Stochastic Programming: A Hierarchy of Bounds and a Partition Sampling*, preprint, http://www.optimization-online.org/DB_HTML/2016/01/5311.html, 2016.

[4] S. P. DOKOV AND D. P. MORTON, *Second-order lower bounds on the expectation of a convex function*, Math. Oper. Res., 30 (2005), pp. 662–677.

[5] N. C. P. EDIRISINGHE, *Bound-based approximations in multistage stochastic programming: Nonanticipativity aggregation*, Ann. Oper. Res., 85 (1999), pp. 103–127.

[6] N. C. P. EDIRISINGHE AND W. T. ZIEMBA, *Bounding the expectation of a saddle function with application to stochastic programming*, Math. Oper. Res., 19 (1994), pp. 314–340.

[7] N. C. P. EDIRISINGHE AND W. T. ZIEMBA, *Tight bounds for stochastic convex programs*, Oper. Res., 40 (1992), pp. 660–677.

[8] YU. ERMOLIEV, A. GAIVORONSKI, AND C. NEDEVA, *Stochastic optimization problems with incomplete information on distribution functions*, SIAM J. Control Optim., 23 (1985), pp. 697–716.

[9] L. F. ESCUDERO, A. GARÍN, M. MERINO, AND G. PÉREZ, *The value of the stochastic solution in multistage problems*, TOP, 15 (2007), pp. 48–64.

[10] K. FRAUENDORFER *Barycentric scenario trees in convex multistage stochastic programming*, Math. Program., 75 (1996), pp. 277–293.

[11] K. FRAUENDORFER AND M. SCHÜRLE, *Multistage stochastic programming: Barycentric approximation*, P. Pardalos and C. A. Floudas, eds., in Encyclopedia of Optimization, Vol. 3, Kluwer Academic Publishers, Dordrecht, 2001, pp. 576–580.

[12] K. FRAUENDORFER, D. KUHN, AND M. SCHÜRLE, *Barycentric bounds in stochastic program-*

ming: Theory and application, in Stochastic Programming: The State of the Art, G. Infanger, ed., Springer, New York, 2011, pp. 67–96.

[13] C. C. Huang, I. Vertinsky, and W. T. Ziemba, *On multiperiod stochastic dominance*, J. Financ. Quant. Anal., 13 (1978), pp. 1–13.

[14] J. L. Jensen, *Sur les fonctions convexes et les inégalités entre les valeurs moyennes*, Acta Math., 30 (1906), pp. 175–193.

[15] D. Kuhn, *Generalized Bounds for Convex Multistage Stochastic Programs*, Lecture Notes in Econom. and Math. Systems 548, Springer, Berlin, 2005.

[16] D. Kuhn, *Aggregation and discretization in multistage stochastic programming*, Math. Program., 113 (2008), pp. 61–94.

[17] H. Levy and J. Paroush, *Multi-period stochastic dominance*, Manag. Sci., 21 (1974), pp. 428–435.

[18] A. Madansky, *Bounds on the expectation of a convex function of a multivariate random variable*, Ann. Math. Stat., 30 (1959), pp. 743–746.

[19] A. Madansky, *Inequalities for stochastic linear programming problems*, Manag. Sci., 6 (1960), pp. 197–204.

[20] F. Maggioni, E. Allevi, and M. Bertocchi, *Bounds in multistage linear stochastic programming*, J. Optim. Theory Appl., 163 (2014), pp. 200–229.

[21] F. Maggioni, E. Allevi, and M. Bertocchi, *Monotonic bounds in multistage mixed-integer linear stochastic programming*, Comput. Manag. Sci., 13 (2016), pp. 423–457.

[22] F. Maggioni and G. Ch. Pflug, *Bounds and approximations for multistage stochastic programs*, SIAM J. Optim., 26 (2016), pp. 831–855.

[23] G. Mokobodski, *Elements extrémeaux pour le balayage*, Séminare Brelot-Choquet-Deny: 13ième année, Tome 5 (1970).

[24] A. Müller and D. Stoyan, *Comparison Methods for Stochastic Models and Risks*, Wiley, New York, 2002.

[25] G. Ch. Pflug, *Version-independence and nested distributions in multistage stochastic optimization*, SIAM J. Optim., 20 (2009), pp. 1406–1420.

[26] G. Ch. Pflug and A. Pichler, *Multistage Stochastic Programming*, Springer Ser. Oper. Res. Financ. Eng., Springer, New York, 2014.

[27] B. Sandikçi, N. Kong, and A. J. Schaefer, *A hierarchy of bounds for stochastic mixed-integer programs*, Math. Program., 138 (2012), pp. 253–272.

[28] B. Sandikçi and O. Y. Özaltin, *A scalable bounding method for multistage stochastic integer programs*, SIAM J. Optim., 27 (2017), pp. 1772–1800.

[29] G. L. Zenarosa, O. A. Prokopyev, and A. J. Schaefer, *Scenario-Tree Decomposition: Bounds for Multistage Stochastic Mixed-Integer Programs*, Working paper, Department of Industrial Engineering, University of Pittsburgh, Pittsburgh, PA, http://www.optimization-online.org/DB_HTML/2014/09/4549.html, 2014.