INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

WILEY

INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

Intl. Trans. in Op. Res. 0 (2023) 1
DOI: 10.1111/itor.13317

On the safe side of stochastic programming: bounds and approximations

Benjamin S. Narum[a], Francesca Maggioni[b] and Stein W. Wallace[a,*]

[a] *Department of Business and Management Science, NHH Norwegian School of Economics, Helleveien 30, 5045, Bergen, Norway*

[b] *Department of Management, Information and Production Engineering, University of Bergamo, Viale G. Marconi 5, Dalmine, BG 24044, Italy*

E-mail: benjamin.narum@nhh.no[Narum]; francesca.maggioni@unibg.it[Maggioni]; stein.wallace@nhh.no[Wallace]

Received 6 April 2023; received in revised form 4 May 2023; accepted 4 May 2023

Abstract

There will always be stochastic programs that are too large or complex to be solved in their basic form. In this article, we review, discuss, and compare different ways such stochastic programs can be handled using bounds and approximations, all based on manipulations of the random variables. We are particularly interested in how methods based on different underlying ideas can be combined or possibly are the same.

Keywords: stochastic programming; bounds; distribution approximation; function approximation

1. Introduction

Stochastic programming is an important tool for dealing with decision problems when the outcome of a decision is stochastic. This is qualitatively different from deterministic approaches and considers concepts such as options, hedging and risk that do not exist in a deterministic world. Optimal decisions to stochastic programs consider different future outcomes simultaneously and consequently allocate resources to obtain flexibility in dealing with these futures effectively if they occur.

The advantage of stochastic programming, as opposed to other approaches to decision-making under uncertainty, is its tight link to mathematical programming and its tools to solve large-scale constrained optimisation problems that may include combinatorial considerations by integers. The field of mathematical programming has seen immense improvements in the ability to get precise solutions to optimisation problems; however, stochastic programs also require *approximation of the stochastics*. As a result, the mathematical program solved may be very close to, or at times very far

* Corresponding author.

from, the underlying problem with its stochastics. It all depends on the quality of the approximation.

Stochastic programming problems are formulated as optimising the expectation (or some other risk measure) over stochastic variables that take the role of parameters in an optimisation problem. Exact solutions to these are, in general, fundamentally intractable.

This paper gives an overview of bounds and approximations with respect to the distributional aspects of stochastic programming. Our primary goal is to give an intuitive overview to provide an understanding of the underlying ideas of these bounds. While there are many associated technical details, we aim to simplify the exposition as much as possible while still conveying the principal ideas, distinctions and connections between different approaches. References are listed throughout for the interested reader to explore in more detail.

We refer readers to the book by King and Wallace (2012) for the modelling aspects of stochastic programming while Kall and Wallace (1994), Birge and Louveaux (2011) and Shapiro et al. (2014) explain the mathematical background, solution procedures and more detailed technical properties. This section discusses the uses of bounds and some of their implications for modelling.

Motivation on bounds. Models in stochastic programming aim to optimise a current decision while additionally accounting for the limitations this puts on future decisions determined once the outcomes of uncertain parameters are revealed. Future decisions reflect reactive actions once more information is known and are usually determined by an optimisation model with no closed-form solution. This means the objective function of the current decision is to optimise the expectation of another optimisation problem. Needless to say, evaluation of this expectation poses a great challenge (especially if the distribution is continuous) and must be approximated in some way. Furthermore, we also aim to optimise this expected value (whose evaluation is intractable) with respect to the current decision. The literature on bounds aims to find tight interval estimates of these quantities. In this paper, we explore bounds both on the *evaluation of expected values* and on the *optimised expected values*. The distinction between these is important because they elicit different kinds of argumentation.

In a mathematical and algorithmic sense, bounds are important to determine if an error gap is sufficiently low so that the problem can be considered solved or that estimates of certain quantities are sufficiently accurate. The former relies on some version of the latter but has different conceptual purposes. Bounds on the evaluation of expectations can be used to reveal the current performance of a given solution in terms of the original (intractable) objective function, but we are also interested in whether the solution could have been better by comparing it to an interval estimate of the optimal objective value. This distinction is especially interesting since finding a solution is subject to (an often different) approximation, and the effectiveness of this approximation can be benchmarked by first estimating its performance and then deciding if this is sufficiently close to being optimal.

There is also an application side of bounds. In the context of minimisation (which we use throughout), a *lower bound is optimistic* while an *upper bound is conservative*. This means we are generally more interested in upper bounds since these are more informative of our true exposure. Investment and pricing problems are especially interesting areas to apply the evaluation bounds in Section 3 because such problems consist of deciding whether to get exposed or not, and evaluating this exposure is essential. An overly optimistic account of the objective value may lead to the

phenomenon known as the *winner's curse* (Thaler, 1988), where the bidder having the worst value estimate wins the bid. If we are already invested, the perspective changes and we mostly care about making good decisions. The bounds in Section 4 are then of particular interest since these can be used to evaluate how close a solution is to the actual optimal objective value.

Generally, more uncertainty leads to a worse performance if we do not plan for it. Ignoring uncertainty produces a decision whose real performance is an upper bound on the best attainable performance and gets better with a better account of uncertainty. This is worth remembering when deterministic models are applied in stochastic environments. A naive approximation of the underlying distribution may also lead to bounds instead of estimates, a surprisingly easily made mistake.

The structure of the paper is as follows: background and notation are described in Section 2. In Section 3, we explore bounds on the evaluation of expected values, while Section 4 emphasises bounds on the optimal value of such expected values. The argumentation of the approaches in these two sections is fundamentally different but has interesting connections. In Section 5, we conclude the paper.

2. Background and notation

We present the notation used throughout the paper for *minimisation* of stochastic programs under uncertainty. Such formulations may be stated as *two-stage* or *multistage* problems, but the same notation is used for both. Unless explicitly stated otherwise, we refer to both kinds of formulations. Risk aversion is omitted, which is why these two formulations are very similar in notation. We assume all relevant optimisation problems in the paper have finite optimal solutions.

We consider a finite-horizon sequential decision-making problem under uncertainty where decisions are made at discrete stages $t \in \{0, \dots, T\}$ and T denotes the (finite) planning horizon, while $[t] := \{0, \dots, t\}$ denotes all stages up to t. The uncertain data $\xi := (\xi_1, \dots, \xi_T)$ are revealed gradually over time from stage $t = 1$, and our decisions should be adapted to this process. The decision process takes the form

$$\underset{\text{decision}}{x_0} \rightarrow \underset{\text{observation}}{\xi_1} \rightarrow \underset{\text{decision}}{x_1} \rightarrow \cdots \rightarrow \underset{\text{observation}}{\xi_T} \rightarrow \underset{\text{decision}}{x_T},$$

where each ξ_t has support $\Xi_t \in \mathbb{R}^{d_t}$, and overall ξ has support $\Xi := \Xi_1 \times \cdots \times \Xi_T$. The history up to stage t is denoted $\xi_{[t]} := (\xi_1, \dots, \xi_t)$. For consistency, we let ξ_0 represent an initial deterministic state known in advance of the first decision. The decision process begins with initial decision x_0 at stage $t = 0$, called the *first-stage* or *here-and-now* decision, followed by sequential decisions x_t at stages $t = 1, \dots, T$. The history of the decision process, at a given point in time t, is denoted by $x_{[t]} := (x_0, x_1, \dots, x_t)$.

An important concept associated with uncertainty is the corresponding *information structure* $\mathcal{F} := (\mathcal{F}_0, \dots, \mathcal{F}_T)$, which describes the availability of information at each stage. We refer to outcomes by the underlying outcome space Ω, and information is encoded by whether such outcomes can be distinguished at a given point in time. Here, $\mathcal{F}_t$ denotes the available information at stage t, and we have the criterion that $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$ which signifies that information is gained over time. Continuous outcome spaces require the notion of nested σ-algebras over Ω to encode the revelation

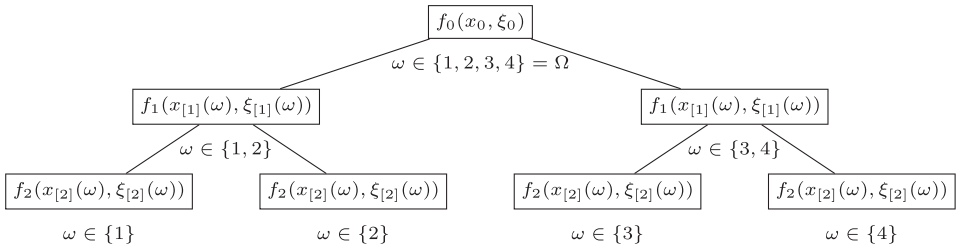

Fig. 1. A scenario tree representing the information structure and the objective value associated with different nodes.

of information, while for discrete outcome spaces, this can also be represented by a tree structure (see Fig. 1). Implicitly, $\xi_t : \Omega \to \mathbb{R}^{d_t}$ is here a mapping from outcomes ω to realisations of the uncertain data $\xi_t(\omega)$, and similarly for the decision process $x_t : \Omega \to \mathbb{R}^{n_t}$. We generally also refer to $x(\omega) = (x_0, x_1(\omega), \ldots, x_T(\omega))$ as a *decision policy*, which is said to be *feasible* if it satisfies the constraints $x(\omega) \in \mathcal{X}(\omega) := \mathcal{X}_0(\omega) \times \cdots \times \mathcal{X}_T(\omega)$, where $\mathcal{X}_t(\omega) \subseteq \mathbb{R}^{n_t}$. Whenever the specific value of the uncertain data process matters, we refer to it by ξ, while we generally use ω as a generic referral to outcomes. Note that processes $\xi(\omega)$ and $x(\omega)$ must always adhere to the information structure $\mathcal{F}$, meaning they are defined with respect to it. This also means the decision policy $x_t(\omega)$ at stage t may only depend on information $\mathcal{F}_t$ available up to stage t, also known as *nonanticipativity*.

We let $f_t(x_{[t]}(\omega), \xi_{[t]}(\omega))$ denote the cost incurred at stage t given the decisions $x_{[t]}(\omega)$ and stochastic process $\xi_{[t]}(\omega)$ up to stage t, given by outcome ω. Let $\mathbb{E}_t^P[\cdot] := \mathbb{E}_{\mathcal{F}_t}^P[\cdot]$ denote the conditional expectation with respect to all information up to stage t using distribution P, and let $\mathbb{E}^P[\cdot]$ denote the unconditional expectation.

A $(T+1)$-stage stochastic programming problem can be formulated as

$$v(P) := \min_{x_0 \in \mathcal{X}_0} f_0(x_0, \xi_0) + \mathbb{E}_0^P\Bigg[\min_{x_1(\omega) \in \mathcal{X}_1(\omega)} f_1(x_{[1]}, \xi_{[1]}) + \mathbb{E}_1^P\Bigg[\cdots + \mathbb{E}_{T-1}^P\Bigg[\min_{x_T(\omega) \in \mathcal{X}_T(\omega)} f_T(x_{[T]}, \xi_{[T]})\Bigg]\Bigg]\Bigg], \tag{1a}$$

$$= \min_{x(\omega) \in \mathcal{X}(\omega)} \mathbb{E}^P\left[\sum_{t=0}^{T} f_t(x_{[t]}(\omega), \xi_{[t]}(\omega))\right] = \min_{x(\omega) \in \mathcal{X}(\omega)} \mathbb{E}^P[f(x(\omega), \xi(\omega))], \tag{1b}$$

where $v(P)$ is the optimal objective value under distribution P and the shorthand notation $f(x(\omega), \xi(\omega))$ in (1b) denotes the total incurred cost for the given process ξ and decision x in outcome ω. Notice that the cost functions f_t can possibly be non-linear.

For discrete outcome spaces Ω, it is useful to depict the principle of information structure as a tree where each ω represents a path from the root to the leaves. This is illustrated in Fig. 1 with the associated notation. The tree has nodes organised in levels corresponding to stages $t \in \{0, \ldots, T\}$. At level $t = 0$, we have a single *root node* associated with the known quantity ξ_0 and the first-stage decision x_0. At level $t = 1$, we have multiple nodes representing different realisations of ξ_1 and associated decisions x_1 that depend on this realisation. Generally, nodes correspond to

possible realisations of uncertainty, and each node has an associated set of outcomes that describes which paths pass through it. Information consists of knowing this set of outcomes but without the ability to distinguish exactly which one occurs in the future. Hence, we must plan for all of them simultaneously (weighted by probability).

By the theory of duality in optimisation, we may dualise this optimisation problem (1) with respect to its constraints $\mathcal{X}(\omega)$, and obtain

$$v(P) = \min_{x(\omega)\in\mathcal{X}(\omega)} \mathbb{E}^P[f(x(\omega),\xi(\omega))] \geq \max_{\lambda(\omega)\in\Lambda(\omega)} \mathbb{E}^P[f^*(\lambda(\omega),\xi(\omega))] = v^*(P), \tag{2}$$

by weak duality, where $f^*(\lambda(\omega),\xi(\omega))$ is the dual objective function, $v^*(P)$ its optimal objective value, $\lambda(\omega)$ a dual policy and $\Lambda(\omega)$ the dual constraints. All quantities in the dual problem also adhere to the same information structure $\mathcal{F}$ and follow the same notation conventions.

2.1. Stochastic linear programs

Stochastic linear programs are the primary formulation used for applications of stochastic programming and may also be extended by including integer variables. These take a particular form with linear objective $f(x(\omega),\xi(\omega)) = c(\omega)^\top x(\omega)$ and linear constraints $\mathcal{X}(\omega) = \{x(\omega) : A(\omega)x(\omega) \geq b(\omega)\}$. Properties of these are used as assumptions in Sections 3.4 and 4.3.3, while its parametric shape is important for the discussion in Section 3.3. The stochastic linear program (SLP) has formulation

$$\begin{aligned} \min_{x(\omega)} \quad & \mathbb{E}^P\big[c(\omega)^\top x(\omega)\big] \\ \text{s.t.} \quad & A(\omega)x(\omega) \geq b(\omega), \quad \forall\omega\in\Omega \\ & x(\omega) \geq 0 \end{aligned} \tag{SLP}$$

with dual

$$\begin{aligned} \max_{\lambda(\omega)} \quad & \mathbb{E}^P\big[b(\omega)^\top \lambda(\omega)\big] \\ & A(\omega)^\top\lambda(\omega) \leq c(\omega), \quad \forall\omega\in\Omega \\ & \lambda(\omega) \geq 0, \end{aligned} \tag{D-SLP}$$

where $\xi(\omega)$ is represented by the random parameters $c(\omega)$, $A(\omega)$ and $b(\omega)$ while $\lambda(\omega)$ denotes the dual multiplier of the constraints of (SLP). This general formulation of the constraints has a time structure such that $\xi_t(\omega)$ denotes the parameters $c_t(\omega)$, $A_t(\omega) := (B_t(\omega), W_t(\omega))$ and $b_t(\omega)$ specific

to stage t. The constraint matrix $A(\omega)$ has a block structure incorporating decisions at previous stages. In particular, the constraints in the primal take the more specific form

$$A(\omega)x(\omega) \geq b(\omega) \iff \begin{cases} A_0 x_0 & \geq b_0, \\ B_1(\omega)x_0 + W_1(\omega)x_1(\omega) & \geq b_1(\omega), \\ B_2(\omega)x_{[1]}(\omega) + W_2(\omega)x_2(\omega) & \geq b_1(\omega), \\ \vdots & \\ B_T(\omega)x_{[T]}(\omega) + W_T(\omega)x_T(\omega) & \geq b_T(\omega), \end{cases}$$

where the first constraint is deterministic, and the consecutive ones are stochastic.

There are some specific shapes and properties of (SLP) to take into consideration:

- *Right-hand side uncertainty*: If only the right-hand side $b(\omega)$ is random, (SLP) is convex and piecewise linear in the random parameters.
- *Objective uncertainty*: If only the objective coefficient $c(\omega)$ is uncertain, (SLP) is concave and piecewise linear in the random parameters.
- *Convex-concave saddle*: If we have both right-hand side $b(\omega)$ and objective $c(\omega)$ uncertainty, (SLP) is piecewise bi-linear in the random parameters and is also a convex-concave saddle function.
- *Fixed recourse*: If W_t is deterministic, this is referred to as *fixed recourse*, while otherwise this it is called *random recourse*. The shape of (SLP) in W is generally non-convex.
- *Complete recourse*: The assumption that (SLP) is always feasible, i.e., taking finite values, is referred to as complete recourse. Relatively complete recourse means feasibility in stage t is ensured for all previous feasible decisions $\mathcal{X}_{[t-1]}$.

These shapes follow from the parametric shape of LPs (Bertsimas and Tsitsiklis, 1997). Notice that bounds based on convexity are generally not applicable to mixed-integer linear programs .

2.1.1. Penalty formulations and soft-constraints

Penalty formulations of (SLP) using soft-constraints are especially useful to consider in the context of bounds. These can be used to define finite growth conditions (Section 3.3.5), to construct majorising functions (Section 3.4), and to have well-defined behaviour for sub-optimal candidate policies (Section 4). Soft constraints are constructed by adding non-negative auxiliary variables $z(\omega)$ to the constraints as

$$A(\omega)x(\omega) + z(\omega) \geq b(\omega), \tag{3}$$

that have associated penalties $\mu \geq 0$ in the objective. This special form of constraint allows us to determine the solution of z easily as

$$z(\omega) = (b(\omega) - A(\omega)x(\omega))^+ := \max\{0, b(\omega) - A(\omega)x(\omega)\}, \tag{4}$$

whose form is similar to those of simple recourse problems (Birge and Louveaux, 2011, section 3.1d). Observe also that this reduces projection onto the feasible set to a trivial max operation. The primal objective can then be restated as

$$c(\omega)^\top x(\omega) + \mu^\top (b(\omega) - A(\omega)x(\omega))^+, \tag{5}$$

while relaxing all constraints but the variable bounds on $x(\omega)$. A direct consequence of soft constraints is that we have bounds $0 \leq \lambda(\omega) \leq \mu$ on the dual variables. Intuitively, whenever the dual multiplier (the shadow price) reaches μ, the penalty takes over instead by violating the constraint. Consequently, penalties should be chosen according to the true marginal cost of violating these constraints. Analogously, soft constraints in the dual formulation using penalties ν give bounds $0 \leq x(\omega) \leq \nu$ on the primal variable, and the dual formulation is always feasible. The dual objective can then be restated as

$$b(\omega)^\top \lambda(\omega) + \nu^\top \big(A(\omega)^\top \lambda(\omega) - c(\omega)\big)^+, \tag{6}$$

while relaxing all constraints but the variable bounds on $\lambda(\omega)$.

The geometric interpretation of these penalty formulations is that the optimal objective values of the primal and dual formulations, as functions of the stochastic parameters, are enclosed in pointed cones whose growth rate is determined by the penalties μ and ν. In the special case of simple recourse, the objective function is equal to this cone.

3. Bounds on the evaluation of expectations

When the underlying distribution P is known, for example, by an analytical expression or by specific information about its moments, there exist methods to get guaranteed (as opposed to statistical) bounds on the expected objective value without solving the intractable integral implied by the original problem formulation (1). An essential property of the bounds reviewed in this section is that they require evaluating the objective function only on a finite set of outcomes, also when the underlying distribution is continuous.

In this section, we primarily address two-stage problems (i.e., $T = 1$) for a *fixed* first-stage decision x_0 and uncertainty in the form of a *random vector* ξ on support $\Xi \subseteq \mathbb{R}^d$. For ease of notation in this section (Section 3), we let ξ_i denote component i of the random vector and omit the time index of ξ. The function to be approximated is then stated as

$$\phi(\xi) := \xi \mapsto \min_{x_1 \in \mathcal{X}_1(\omega)} f_1(x_{[1]}, \xi), \qquad x_0 \text{ fixed}, \tag{7}$$

which represents the optimisation problem solved once the uncertainty ξ is revealed. This can be time-consuming to evaluate, which motivates having few evaluation points to determine bounds on its expectation $\mathbb{E}^P[\phi(\xi)]$. Extension to multistage problems is discussed in Section 3.3.6.

Guaranteed bounds are mainly based on the shape of ϕ. These bounds have two closely linked interpretations, but many of the bounds we discuss have been developed primarily with one perspective in mind. The two interpretations of guaranteed bounds are

1. **Distribution approximation**: Find an alternative distribution Q whose evaluation of the expected objective function gives a bound.
2. **Function approximation**: Find a simpler function that is strictly equal or larger/smaller than ϕ and whose expectation is easily evaluated.

Distribution approximation can intuitively be understood as moving probability mass to parts of the distribution support that combined has a larger/smaller objective value. Function approximation is intuitively understood by the fact that a function that is larger than another must also have a larger expectation.

The link between distribution approximation and function approximation and the more general framework of generalised moment problems are discussed in Section 3.1. An interesting generalising property of distributional bounds is discussed in Section 3.2. Bounding approaches primarily motivated by distribution approximation are discussed in Section 3.3 and those motivated by function approximations in Section 3.4. Lastly, Monte Carlo integration is an often used alternative approach to evaluate expectations, discussed in Section 3.5.

3.1. Guaranteed bounds as a generalised moment problem

The *generalised moment problem* (GMP) has important implications for deriving bounds on the expectation of a function when certain limited information about the distribution is known. To some degree, this underlies all of the guaranteed bounds. The GMP is a semi-infinite program whose formulation aims to optimise the expectation of a function over an altered distribution that is subject to certain generalised moment constraints. While its formulation is very general and solutions are not always easily given, certain setups of the GMP give easily found or even analytical solutions. We primarily explore these easily found solutions in this paper and leave out details about semi-infinite programs.

The link to bounds in stochastic programming started with Madansky (1959), who used the moment problem to prove upper bounds on convex functions and was later tied to optimisation of the distribution to get upper and lower bounds by Dupačová (1987), Klein Haneveld (1986) and Kall (1988), originally pioneered by Dupačová (1966) with the minimax formulation of stochastic programs. The optimal distribution of the GMP has also been referred to as an extreme measure which is particularly interesting because it has been shown that these are discrete and finite, even if the original distribution P is continuous (Kemperman, 1968; Karr, 1983). Duality further shows that the distribution approximation by the GMP has dual formulation with the direct interpretation of function approximation, which makes these two perspectives particularly tightly connected. The GMP can be solved directly to obtain bounds but is based on an iterative procedure that involves solving non-convex sub-problems (Birge and Wets, 1987), and we instead emphasise simpler procedures.

We now state the mathematical framework of guaranteed bounds starting at the interpretation of distribution approximation, function approximation and implications of their link by strong duality. Lastly, we give an illustrative example in Section 3.1.1.

Distribution approximation. Consider that we want to find an alternative distribution Q on support Ξ to evaluate the expectation $\mathbb{E}^Q[\phi(\xi)]$ such that it bounds the true objective $\mathbb{E}^P[\phi(\xi)]$ from above or below. The first observation to make is that the expectation is linear in the distribution (namely, in the probability assigned to each outcome) irrespective of the functional shape of $\phi(\xi)$. This means we can pose the bounding problem as a (semi-infinite) linear program where the distribution Q is optimised with respect to the expectation $Q \mapsto \mathbb{E}^Q[\phi(\xi)]$, subject to constraints $Q \in \mathcal{P}$ where $\mathcal{P}$ is a specific class of probability distributions. We may then intuitively maximise to get an upper bound and minimise to get a lower bound. The formulation for obtaining a lower bound by distribution alteration (thereof the abbreviation D) is

$$\min_{Q \in \mathcal{P}} \mathbb{E}^Q[\phi(\xi)], \tag{D-LB}$$

and

$$\max_{Q \in \mathcal{P}} \mathbb{E}^Q[\phi(\xi)], \tag{D-UB}$$

for the upper bound. An optimal solution to such formulations is denoted Q^*. In general, we must assume the support Ξ is bounded for Q^* to be well defined (extensions to unbounded support are discussed in Section 3.3.5).

Conveniently, the optimisation problems (D-LB) and (D-UB) are optimal at discrete distributions Q^*, even if P is continuous. This is given that we have a finite number of constraints of the general form

$$\mathbb{E}^Q[g_i(\xi)] = m_i := \mathbb{E}^P[g_i(\xi)], \quad \forall i \in \mathcal{I}, \tag{8}$$

where $\mathcal{I}$ is a finite index set of constraints (Kemperman, 1968, Theorem 1). The *generalised moment functions* $g_i(\xi)$ for $i \in \mathcal{I}$ can take various forms but should be linearly independent. Common expressions for these are linear functions, moment functions or multi-linear functions. We also require that probabilities under Q sum to one, which can be stated in the same form,

$$\mathbb{E}^Q[1] = Q(\Xi) = 1 = P(\Xi) = \mathbb{E}^P[1].$$

The constraint set on Q is then summarised as

$$\mathcal{P} := \left\{ Q : \begin{array}{c} Q(\Xi) = 1, \\ \mathbb{E}^Q[g_i(\xi)] = m_i, \quad \forall i \in \mathcal{I} \end{array} \right\}. \tag{9}$$

We assume that the prescribed values m_i are easily found and that they are consistent with each other.[1] According to the definition of m_i. we always have that $P \in \mathcal{P}$, which means P is always feasible to (D-LB) and (D-UB). Their respective optimal solutions Q^* must thus indeed provide bounds on $\mathbb{E}^P[\phi(\xi)]$ by optimising over the same set $\mathcal{P}$. Furthermore, the optimal approximating distribution Q^* has at most $|\mathcal{I}| + 1$ points in its support (Kemperman, 1968). The contributions on different bounds in the literature lie in choosing generalised moment functions $g_i(\xi)$ such that Q^* is easily found, which is treated in more detail in Section 3.3.

[1] If freely specified, it is possible to choose m_i so that $\mathcal{P}$ is empty.

Function approximation. We now turn to the perspective of function approximation. Instead of looking for a new distribution Q to evaluate expectations over basis functions $g_i(\xi)$, we now aim to find an affine combination of them

$$\tilde{\phi}(\xi) := u_0 + \sum_{i \in \mathcal{I}} u_i g_i(\xi), \tag{10}$$

that approximates ϕ well from either above or below, where u is a vector of $|\mathcal{I}| + 1$ elements that must be chosen appropriately. Recall that the true expectations of $g_i(\xi)$ are known to be m_i, so

$$\mathbb{E}^P\big[\tilde{\phi}(\xi)\big] = u_0 + \sum_{i \in \mathcal{I}} u_i m_i. \tag{11}$$

In the case of a lower bound, we constrain u such that $\tilde{\phi}(\xi)$ approximates $\phi(\xi)$ from below, while for the upper bound, u is constrained such that $\tilde{\phi}(\xi)$ approximates $\phi(\xi)$ from above. The approximating problem from the perspective of function approximation (thereof the abbreviations F) has the formulation

$$\max_u \left\{ \mathbb{E}^P\big[\tilde{\phi}(\xi)\big] = u_0 + \sum_{i \in \mathcal{I}} u_i m_i \quad \text{s.t.} \quad \tilde{\phi}(\xi) \le \phi(\xi), \quad \forall \xi \in \Xi \right\}, \tag{F-LB}$$

for a lower bound and

$$\min_u \left\{ \mathbb{E}^P\big[\tilde{\phi}(\xi)\big] = u_0 + \sum_{i \in \mathcal{I}} u_i m_i \quad \text{s.t.} \quad \tilde{\phi}(\xi) \ge \phi(\xi), \quad \forall \xi \in \Xi \right\}, \tag{F-UB}$$

for an upper bound. An optimal solution to these formulations is denoted u^*.

Duality. Distribution approximation and function approximation are really two sides of the same coin: they are linked by *strong duality*. This means the distribution alteration problem (D-LB) has dual formulation (F-LB), and these have respective solutions Q^* and u^* whose objective values are equal. Correspondingly for (D-UB) and (F-UB). Strong duality holds by the insightful property that the approximating function $\tilde{\phi}(\xi)$ is *equal* to $\phi(\xi)$ at values of ξ where Q assigns probability mass. This follows from complementarity slackness of the primal-dual pair, which states that

$$Q\big(\tilde{\phi}(\xi) = \phi(\xi)\big) = 1, \tag{12}$$

namely, the probability under Q that the approximating function $\tilde{\phi}(\xi)$ is equal to $\phi(\xi)$ is one. Intuitively, this means the optimal function approximation meets the true function in the support of the optimal approximating distribution.

3.1.1. Example: One-dimensional bounded support

We present an example to illustrate these results on distribution alteration and function approximation to obtain bounds (see Fig. 2). Assume we have a one-dimensional uniform distribution P

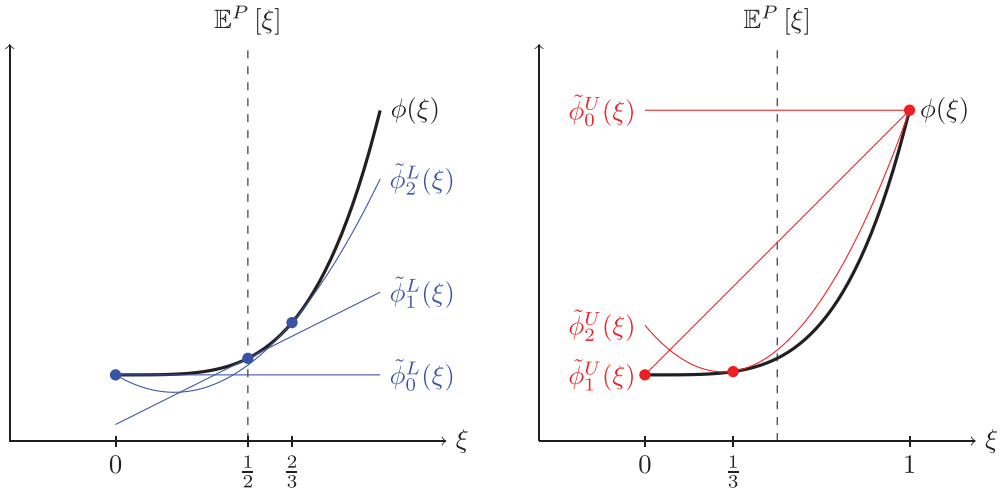

Fig. 2. Examples of lower (left) and upper (right) approximations, with points at the position of the discrete distribution Q support where the approximations meet $\phi(\xi)$.

on support $\Xi = [0, 1]$ with expectation 0.5. Given this distribution, we want to solve (D-LB) and (D-UB) where

$$\phi(\xi) = \xi^4, \qquad g_i(\xi) = \xi^i, \quad \forall i \in \{0, 1, 2\},$$

are the objective and (generalised) moment functions, respectively, where i are the different moments we want to be consistent in the approximating distribution Q. The lower bounding problem is then formulated as

$$\min_Q \left\{ \mathbb{E}^Q[\phi(\xi)] \quad \text{s.t.} \quad \mathbb{E}^Q\left[\xi^i\right] = \mathbb{E}^P\left[\xi^i\right], \quad \forall i \in \{0, 1, 2\} \right\}, \tag{13}$$

and maximising gives an upper bound instead. Denoting with u_i the dual variable of each constraint, the corresponding approximating functions are

$$\tilde{\phi}_M^{L/U}(\xi) = \sum_{i=0}^{M} u_i \xi^i.$$

where $M + 1$ denotes the number of moments enforced, (L) refers to lower and (U) upper bounding functions, respectively. When solving the problem, we incrementally add each moment constraint to illustrate their differences. The lower bounding approximating functions have been found to have expressions

$$\tilde{\phi}_0^L(\xi) = 0, \qquad \tilde{\phi}_1^L(\xi) = -\frac{3}{16} + \frac{1}{2}\xi, \qquad \tilde{\phi}_2^L(\xi) = -\frac{16}{27}\xi + \frac{4}{3}\xi^2,$$

where the approximating distribution supports are $\{0\}, \{1/2\}, \{0, 2/3\}$, respectively. The upper bounding approximating functions have expressions

$$\tilde{\phi}_0^U(\xi) = 1, \qquad \tilde{\phi}_1^U(\xi) = \xi, \qquad \tilde{\phi}_2^U(\xi) = \frac{3}{16} - \frac{19}{16}\xi + 2\xi^2,$$

where the approximating distribution supports are $\{1\}, \{0, 1\}, \{1/3, 1\}$, respectively. These are plotted in Fig. 2.

Observe (in Fig. 2) that the zeroth moment approximation places all probability mass on the lowest and highest points of $\phi(\xi)$ and the corresponding function approximation $\tilde{\phi}$ is constant. This is an intuitive way of making conservative bounds but requires knowledge of the curve's lowest or highest point. When adding the first-moment basis function, we see that the lower bound places all probability mass in the expectation of ξ (not a coincidence) while the upper bound places all probability mass at the ends of the interval support of ξ. These are recoveries of the Jensen lower bound and the Edmund–Mandansky upper bound discussed in Section 3.3. We have also added the second-moment basis function as an illustration that this generalises for further basis functions. However, the support of the optimal Q^* is generally not as easily found, and there are fewer practical bounds based on this kind of information.

3.2. Generalisation to every decision and stochastic dominance

An important additional property of some of the bounds discussed in this section is that the approximating distribution Q may allow to bound $\mathbb{E}^P[f(x, \xi)]$ for *any* decision x. When this is the case, we may *solve* the stochastic program in terms of the objective $\mathbb{E}^Q[f(x, \xi)]$ to obtain a tighter upper bound on the optimal value $v(P)$ than possible when using only a single sub-optimal candidate solution.[2] Formally, this requires finding a Q such that

$$\mathbb{E}^P[f(x, \xi)] \leq \mathbb{E}^Q[f(x, \xi)], \quad \forall x \in \mathcal{X}, \tag{14}$$

for an upper bound (or reversing the inequality for a lower bound). The relevance of this is limited to situations where a specific property of $\xi \mapsto f(x, \xi)$ for all $x \in \mathcal{X}$ is in itself sufficient to obtain the bounding distribution Q without explicit knowledge of the function. The property (14) is referred to as *stochastic dominance* (a wide area of research in itself), but for now, the most relevant property of $\xi \mapsto f(x, \xi)$ is convexity, under which (14) is referred to as *convex stochastic dominance*. Under convexity, we may establish to a great extent where the approximating distribution Q should reside in its support Ξ without further information on the function itself. Finding Q then comes down to properties of the support Ξ, multivariate dependence in P, and the choice of generalised moment functions $g(\xi)$. Overall, the property of convex stochastic dominance greatly simplifies finding solutions to the bounding problems (D-UB) and (D-LB).

Majorising probability measures by ordering relations was explored first by Birge and Wets (1986) and pointed out by Wets (1984). Edirisinghe and Ziemba (1992) used the property of convex

[2]More generally, the aim of minimising $\max_{Q \in \mathcal{P}} \mathbb{E}^Q[f(x, \xi)]$ in terms of x is referred to as Distributionally Robust Optimisation (DRO) in the literature.

dominance to get optimised upper bounds, while Frauendorfer (1996), Kuhn (2005) and Frauendorfer et al. (2011) apply this to convex-concave multistage problems with linear time dependence. Maggioni and Pflug (2019) apply convex stochastic dominance to bound multistage problems with more general stochastic processes and risk-averse objective functions.

3.3. Distribution approximations

This section discusses approaches that are primarily motivated by distribution approximation to find bounds, where some are also solutions to the GMP. The bounds in this section make assumptions on the shape of ϕ, multivariate dependence in P, the kind of support Ξ and which generalised moment functions $g(\xi)$ are used in the approximation. Table 1 gives an overview of different approaches.

There are four relevant shapes of ϕ to consider: (i) convex, (ii) concave, (iii) convex-concave and (iv) monotonous. Convexity serves as a base case for our analysis. If it is instead concave, the analysis is the same, but lower and upper bounds must be switched. If it is instead convex-concave, these must be dealt with in an opposing manner, but the analysis is still largely the same (Section 3.3.4). Monotonous functions can be dealt with (although less effectively) by finding the highest or lowest evaluations in the support and may be considered a special case of the GMP using only zeroth-order conditions (see Example 3.1.1).

3.3.1. Jensen, Edmund–Madansky and hyper-planes on simplicial support

Jensen (1906) proved the (now classic) result that, for a convex function ϕ of a random variable ξ, its expectation $\mathbb{E}^P[\phi(\xi)]$ is bounded from below by the evaluation in the expectation of the random variable:

$$\phi(\mathbb{E}^P[\xi]) \leq \mathbb{E}^P[\phi(\xi)]. \tag{15}$$

Furthermore, this only requires evaluation of ϕ in a single point, which is of great practical importance when function evaluations can be time-consuming.

Edmundson (1957) showed how the expectation of a convex function of a one-dimensional random variable with a bounded interval support could be upper bounded by a linear function going through the end-points of the interval. Madansky (1959) posed this in the form of moment problems and showed it could be applied repeatedly for multivariate distributions with independence on rectangular support. Frauendorfer (1988) generalised these results to distributions with dependence by instead using multi-linear (generalised) moment functions

$$g_{\mathcal{I}}(\xi) = \prod_{i \in \mathcal{I}} \xi_i, \quad \forall \mathcal{I} \subseteq \{1, \ldots, d\}, \tag{16}$$

where $\mathcal{I}$ are all combinations of each of the dependent random variables (i.e., in the power-set of $\{1, \ldots, d\}$). This gives an analytical expression for optimal probability mass to place in the extreme points on the rectangular support, while Kall (1987) showed this solution is unique. Furthermore, it was shown that the Edmund–Madansky result is a special case of the same expression where

Table 1
Distribution approximation bounds

Name	Bound type	Multivariate dependency	Generalised moment function	Support	# supp(Q^*)	Convex dominance	Contributor(s)
Jensen's inequality	Lower	Yes	First-order	Unbounded	1	Yes	Jensen (1906)
Edmund–Madansky	Upper	No	First-order	Rectangular	2^d	Yes	Edmundson (1957); Madansky (1959)
Edmund–Madansky with dependence	Upper	Yes	Multi-linear	Rectangular	2^d	Yes	Frauendorfer (1988); Kall (1987)
Hyper-plane	Upper	Yes	First-order	Polytopal	$\geq d+1$	No	Birge and Wets (1986)
Hyper-plane	Upper	Yes	First-order	Polyhedral cone	$\geq d+1$	No	Gassmann and Ziemba (1986)
Hyper-plane	Upper	Yes	First-order	Simplicial	$d+1$	Yes	Birge and Wets (1986); Gassmann and Ziemba (1986)
Barycentric approximation	Upper	Yes	Barycentric	Polytopal	$\geq d+1$	Yes	Birge and Wets (1986); Frauendorfer (1992)
Convex-concave saddle	Upper/ lower	No	First- and cross-order	Rectangular	2^d	Yes	Frauendorfer (1989, 1992)
Convex-concave saddle	Upper/ lower	Yes	First- and cross-order	Simplicial	$d+1$	Yes	Frauendorfer (1989, 1992)
Convex-concave saddle	Upper/ lower	Yes	First- and cross-order	Polyhedral cone	$\geq d+1$	No	Edirisinghe and Ziemba (1994b, 1994a, 1996)
Second-order lower bound	Lower	Yes	Second-order	Simplicial	$d+1$	Yes	Edirisinghe (1996)
Second-order lower bound	Lower	Yes	Second-order	Rectangular	2^d	Yes	Edirisinghe (1996)
Second-order lower bound	Lower	No	Second-order	Rectangular	$\geq 2^d$	Yes	Dokov and Morton (2005)
Higher order bound	Upper	Yes	nth-order	Rectangular	$\binom{d+n}{n}$	Yes	Dokov and Morton (2002)

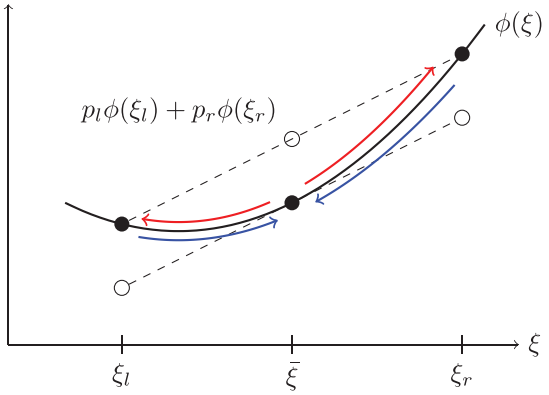

Fig. 3. Moving probability mass on convex functions to obtain bounds with respect to linear approximating functions. Dashed lines illustrate the corresponding function approximations.

independence simplifies the expression. A disadvantage of the Edmund–Madansky (EM) type upper bound is that the number of evaluation points for the altered distribution scales exponentially as 2^d in the dimensionality d of Ξ. Furthermore, it is often unrealistic to have accurate estimates of the expectations of the multi-linear generalised moment functions (16) in real-world settings.

These limitations of the EM type upper bound were mitigated by Birge and Wets (1986) and Gassmann and Ziemba (1986) who, effectively, used first-order moment functions

$$g_i(\xi) = \xi_i, \quad \forall i \in \{1, \ldots, d\}, \tag{17}$$

assuming simplicial support Ξ and general dependence structure. Birge and Wets (1986, Corollary 6.16) show that, under the assumption of convex $\phi(\xi)$ and polytopal support Ξ the bounding problem (D-UB) attains a solution on the extreme points of the polytope. When the polytope is *simplicial* (defined to have $d + 1$ extreme points), the weights on the extreme points are uniquely determined by the set of d equations $\mathbb{E}^P[\xi_i] = \mathbb{E}^Q[\xi_i]$ for all $i \in \{1, \ldots, d\}$, as well as the conditions that probabilities sum to one. From the perspective of function approximation, this corresponds to finding an affine hyper-plane that meets $\phi(\xi)$ in the extreme points of Ξ. Although this makes finding the upper bounding distribution approximation much simpler, the EM type bound is still tighter (Kall, 1987).

Consider also that a bounded interval $[a, b]$ is a simplicial set in one dimension, which means the EM bound is the one-dimensional counterpart of the hyper-plane approach. Independence and rectangular support mean the probabilities can be determined in each dimension separately, and the overall probability in each rectangular corner is determined as the product of the probability in each dimension.

The bounds based on zeroth- and first-order moments can intuitively be understood by directly applying the definition of convexity (see Fig. 3). An affine line can be defined as a convex combination of two points. Placing these points closer to the expectation gives lower evaluations, and placing them further out gives higher evaluations. This observation is generalised by the concept of barycentric approximation for upper bounds.

3.3.2. Barycentric approximation

More generally, we may state upper bounds as making convex combinations of the extreme points of a polytopal support (that is not simplicial) since the probability mass from upper bounding distribution approximations of convex functions always resides at the extreme points of the support (Birge and Wets, 1986). This representation also goes under the name of *barycentric approximation*, referring to the interpretation that a convex combination of the weighted extreme points is a barycentre (centre of mass). These developments are due to Birge and Wets (1986) and Frauendorfer (1992).

First, we express any point ξ within a polytopal set Ξ as a convex combination of its extreme points $\{\xi^{(1)}, \ldots, \xi^{(K)}\} = \text{ext}(\Xi)$ with weights $\{\rho_k(\xi)\}_{k=1,\ldots,K}$, subject to

$$\xi = \sum_{k=1}^{K} \rho_k(\xi)\xi^{(k)}, \qquad \sum_{k=1}^{K} \rho_k(\xi) = 1, \qquad \rho_k(\xi) \geq 0. \tag{18}$$

This is also called a *barycentric coordinate* representation of ξ. The approximating function is then expressed as

$$\tilde{\phi}(\xi) = \sum_{k=1}^{K} \rho_k(\xi)\phi(\xi^{(k)}), \tag{19}$$

which means

$$\phi(\xi) = \phi\left(\sum_{k=1}^{K} \rho_k(\xi)\xi^{(k)}\right) \leq \sum_{k=1}^{K} \rho_k(\xi)\phi(\xi^{(k)}) = \tilde{\phi}(\xi), \tag{20}$$

by the definition that ϕ is convex. Evaluating the expectation on both sides gives the upper bound

$$\mathbb{E}^P[\phi(\xi)] \leq \sum_{k=1}^{K} \bar{\rho}_k\phi(\xi^{(k)}), \tag{21}$$

where $\bar{\rho}_k = \mathbb{E}^P[\rho_k(\xi)]$ is the overall weight assigned to each extreme point given by its weight function $\rho_k(\xi)$. The possible caveat of barycentric approximation is obtaining $\bar{\rho}_k$ since this involves a high-dimensional integral over general functions $\rho_k(\xi)$. Preferably these are stated in a simple fashion. In particular, if Ξ is simplicial, we have that $\rho_k(\xi)$ must be linear and $\bar{\rho}_k$ are uniquely determined, which recovers the hyper-plane result in the previous section. If the support is rectangular, we can recover the dependent EM bound where $\rho_k(\xi)$ are multi-linear and $\bar{\rho}_k$ are again given by closed-form expressions.

3.3.3. Higher order bounds

Edirisinghe and Ziemba (1994b) and Edirisinghe (1996) extend the previous developments by a method using minorising affine functions to obtain lower bounds on convex functions. Originally, these were developed for convex-concave saddle functions (Section 3.3.4) where linearisation in the convex components complements a barycentric approximation in the concave components but also

led to new results on second-order lower bounds on convex functions. This lower bound is the first to improve on the Jensen bound by incorporating additional moment information.

Assume we have a given barycentric approximation $\rho_k(\xi)$ on polyhedral support Ξ, potentially based on a special kind of support (simplicial or rectangular).[3] Since ϕ is convex, we may create a minorising linearisation

$$\phi(\xi) \geq \phi(\hat{\xi}) + \nabla_{\xi}\phi(\hat{\xi})^{\top}(\xi - \hat{\xi}) \tag{22}$$

where $\nabla_{\xi}\phi(\hat{\xi})$ is a sub-gradient of ϕ with respect to ξ at a point $\hat{\xi}$ (where the inequality is also tight). Consider that we may create such a linearisation in certain points $\hat{\xi}^{(k)}$ associated with each function $\rho_k(\xi)$. By the definition of barycentric coordinates, we have that $1 = \sum_{k=1}^{K} \rho_k(\xi)$, which gives

$$\phi(\xi) = \sum_{k=1}^{K} \rho_k(\xi)\phi(\xi) \geq \sum_{k=1}^{K} \rho_k(\xi)\Big[\phi(\hat{\xi}^{(k)}) + \nabla_{\xi}\phi(\hat{\xi}^{(k)})^{\top}(\xi - \hat{\xi}^{(k)})\Big] = \tilde{\phi}(\xi), \tag{23}$$

where the linearisation is specific to each k. Higher order and cross-moment terms may be contained in the term $\rho_k(\xi)\xi$. By evaluating the expectation over (23), we find that

$$\mathbb{E}^{P}[\phi(\xi)] \geq \sum_{k=1}^{K} \bar{\rho}_k\left[\phi(\hat{\xi}^{(k)}) + \nabla_{\xi}\phi(\hat{\xi}^{(k)})^{\top}\left(\frac{\mathbb{E}^{P}[\rho_k(\xi)\xi]}{\bar{\rho}_k} - \hat{\xi}^{(k)}\right)\right] = \sum_{k=1}^{K} \bar{\rho}_k\phi(\hat{\xi}^{(k)}), \tag{24}$$

where the last equality follows by letting $\hat{\xi}^{(k)} = \mathbb{E}^{P}[\rho_k(\xi)\xi]/\bar{\rho}_k$, the point at which the k-specific linearisation of ϕ is tight. Separability of $\tilde{\phi}$ is important as this allows finding each $\hat{\xi}^{(k)}$ separately. The approximating distribution then has support $\{\hat{\xi}^{(k)}\}_{k=1,\ldots,K}$ with probabilities $\hat{\rho}_k$. An illustration for the one-dimensional case is given in Fig. 4.

The points $\hat{\xi}^{(k)}$ here have the interpretation of being the conditional expectation of ξ under the probability assignment $\rho_k(\xi)/\bar{\rho}_k$, which tends towards $\xi^{(k)}$. It also holds that $\hat{\xi}^{(k)} \in \Xi$ as long as $\bar{\rho}_k > 0$ while, otherwise, $\hat{\xi}^{(k)}$ is irrelevant since $\bar{\rho}_k$ is its associated probability mass. If $\rho_k(\xi)\xi$ only contain cross moments, the bound coincides with the solution of (D-LB), while this is not the case for second-order bounds since the approximating distribution only preserves first-order and cross moments exactly. Figure 4 illustrates that the bounding function $\tilde{\phi}(\xi)$ does not meet $\phi(\xi)$, which would be a requirement for this to hold. Dokov and Morton (2005) also provide a tighter second-order lower bound than (23), which requires more evaluation points, rectangular support and independence, whereas Edirisinghe (1996) bound is a special case. Neither this bound solves (D-LB) exactly.

Lastly, Dokov and Morton (2002) extend upper bounds on convex functions to arbitrary order using Bernstein polynomials, which gives a regular grid of evaluation points. As a special case, this also recovers the EM kind of bounds, with and without dependence.

[3]If $\rho_k(\xi)$ are degenerate (i.e., there exist multiple valid representations by the given moment information), the issue of finding the lower bound can instead be formulated as a non-linear convex optimisation problem (Edirisinghe and Ziemba, 1994b, 1996).

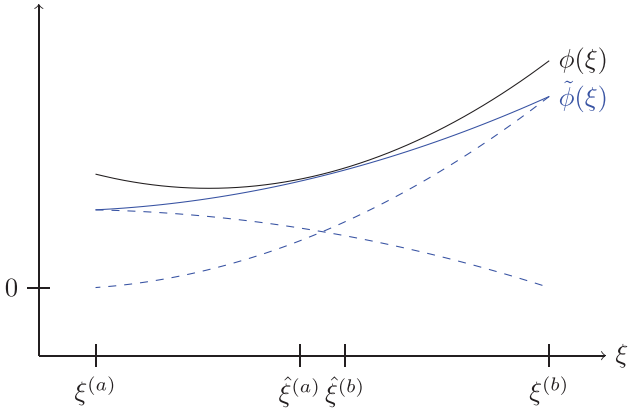

Fig. 4. Second-order lower bound by combined linearisation and barycentric approximation on an interval $[\xi^{(a)}, \xi^{(b)}]$. The evaluation points of the approximating distribution are $\hat{\xi}^{(a)}$ and $\hat{\xi}^{(a)}$, while $\tilde{\phi}$ is the separable approximating function.

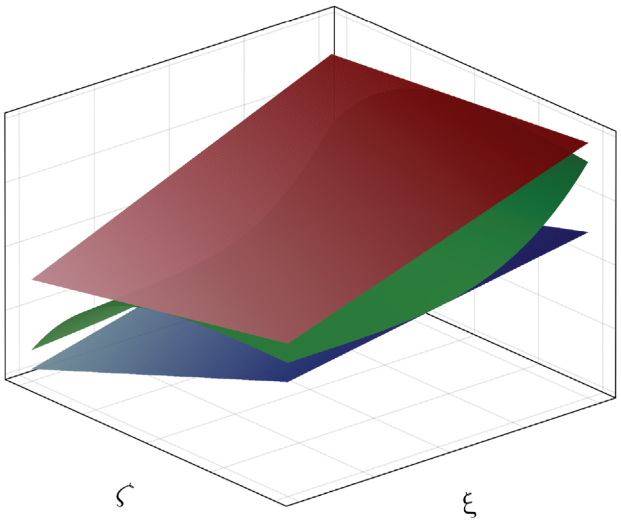

Fig. 5. Bi-linear upper and lower bounding functions on a convex-concave function $\phi(\xi, \zeta)$.

3.3.4. Convex-concave saddle functions

Convex-concave saddle functions, defined to be convex in some dimensions and concave in others, can be dealt with by treating these in an opposing manner. The primary concern for these is how to deal with dependence between the convex and concave components. This section follows the exact developments of Sections 3.3.2 and 3.3.3, while simplifying special cases are due to Frauendorfer (1992). From the perspective of function approximation, these bounds give bi-linear or multi-linear approximating functions on the convex-concave saddle function, as illustrated in Fig. 5.

Let $\phi(\xi, \zeta)$ be a convex-concave saddle function that is convex in ξ and concave in ζ, and let the polytopal distribution support be $\Xi \times Z$ where $\xi \in \Xi$ and $\zeta \in Z$. This corresponds to

appending ζ to the multivariate random variable in the previous setup. Let $\xi^{(k)}$ denote the extreme points of Ξ and $\zeta^{(l)}$ the extreme points of Z. Again, we use barycentric approximations $\rho_k(\xi)$ and $\rho_l(\zeta)$ determined by these extreme points. By the saddle property of $\phi(\xi, \zeta)$ and the barycentric approximation, we then have that

$$\sum_{l=1}^{L} \rho_l(\zeta)\phi(\xi, \zeta^{(l)}) \leq \phi(\xi, \zeta) \leq \sum_{k=1}^{K} \rho_k(\xi)\phi(\xi^{(k)}, \zeta), \tag{25}$$

due to convexity in ξ and concavity in ζ. Evaluating the expectation of (25) gives bounds on $\mathbb{E}^P[\phi(\xi, \zeta)]$; however, this cannot necessarily be easily evaluated and dependence between ξ and ζ must be cared for when doing this. Under independence between ξ and ζ, we may apply the Jensen inequality directly to (25) to obtain the bounds

$$\sum_{l=1}^{L} \bar{\rho}_l\phi_l(\bar{\xi}, \zeta^{(l)}) \leq \mathbb{E}^P[\phi(\xi, \zeta)] \leq \sum_{k=1}^{L} \bar{\rho}_k\phi(\xi^{(k)}, \bar{\zeta}), \tag{26}$$

where $\bar{\xi} = \mathbb{E}^P[\xi]$ and $\bar{\zeta} = \mathbb{E}^P[\zeta]$. Under dependence, however, we may proceed by applying the linearisation procedure from Section 3.3.3 to expression (25) and obtain

$$\sum_{kl} \bar{\rho}_l\phi_l(\hat{\xi}^{(k)}, \zeta^{(l)}) \leq \mathbb{E}^P[\phi(\xi, \zeta)] \leq \sum_{kl} \bar{\rho}_k\phi(\xi^{(k)}, \hat{\zeta}^{(l)}), \tag{27}$$

instead, where $\hat{\xi}^{(k)}$ and $\hat{\zeta}^{(l)}$ are determined by the procedures laid out in Section 3.3.3. Note that the linearisation with respect to ξ uses the barycentric approximation $\rho_l(\zeta)$ (as well as the converse), meaning the approximating functions have terms $\rho_k(\xi)\zeta$ and $\rho_l(\zeta)\xi$ expressed by first-order or cross moments only. As before, the bound using only first-order and cross moments are tight on the corresponding generalised moment problems (D-LB) and (D-UB). Under assumptions of simplicial or rectangular support, the barycentric approximations are unique, and the solutions are easily found in closed form as well as for the linearisation points (Frauendorfer, 1992). Second-order moments can be added by using a single barycentric approximation defined in terms of the extreme points of $\Xi \times Z$ instead (Edirisinghe, 1996).

3.3.5. Subset refinement and unbounded support

The described bounds can be sharpened to arbitrary precision by refining the support Ξ into smaller subsets $\{\Xi_k\}_{k=1,\ldots,K}$ and computing the bounds on each of them. In the perspective of distribution alteration, we now use conditional expectations instead of total expectations while from the perspective of function approximation, we use piece-wise approximations within each subset instead of on the whole support.

The observation that these bounds could be refined to subsets started with Ben-Tal and Hochman (1972) who refined the Edmund–Madansky bound for positive linear functions, i.e. $(\xi)^+ = \max\{0, \xi\}$, and the implications of their result were generalised by Huang et al. (1977) to show that the bounds can be repeatedly applied to subsets of the support to refine them to arbitrary precision. Ben-Tal and Hochman (1972) showed that the bounds hold on intervals unbounded on

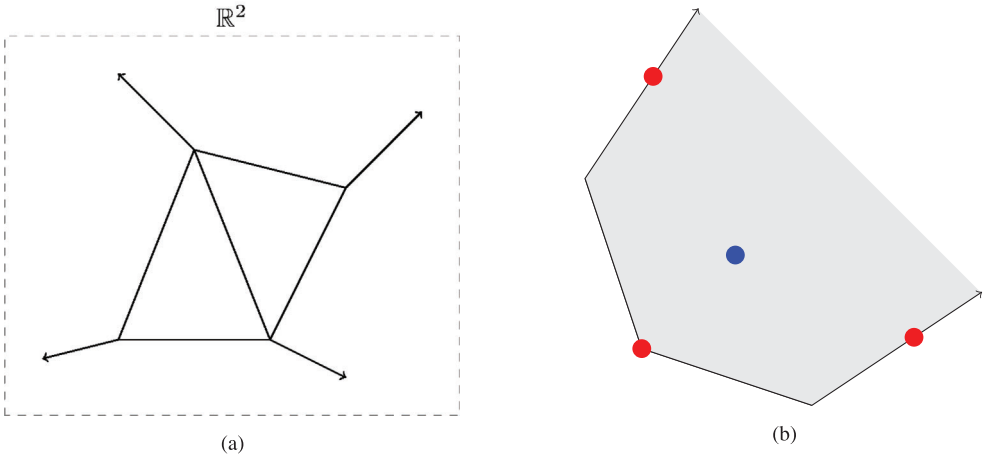

Fig. 6. (a) Partitioning a two-dimensional plane into polytopes and polyhedral cones. (b) Polyhedral cone with corresponding distribution approximations for an upper bound at extreme points extended by rays and a lower bound in the interior.

one side if the convex function has a finite growth rate, which was extended to conic polyhedral sets by Gassmann and Ziemba (1986) and Birge and Wets (1986). Edirisinghe and Ziemba (1994a) extend the lower bounds by linearisation (Section 3.3.3) to conic polyhedral support.

The subsets Ξ_k of Ξ must generally be polytopal or polyhedral cones, described by a finite number of extreme points and rays (see Fig. 6a) and partition the support Ξ. A *polyhedral cone* is a polyhedron where for every unbounded direction it contains, it is bounded in the opposite direction. This avoids unbounded formulations of (D-LB) and (D-UB). Once a bound $B(\Xi_k)$ is made on a subset Ξ_k, the overall bound $B(\Xi)$ is found as $B(\Xi) = \sum_{k=1}^{K} P(\Xi_k)B(\Xi_k)$, where $P(\Xi_k)$ denotes the probability of each subset.

Unbounded support requires the use of rays in its description. The fundamental requirement for unbounded support is that ϕ has limited growth along any ray r staring at a point $\xi \in \Xi$. Namely,

$$\lim_{t\to\infty} \frac{|\phi(\xi + tr)|}{\|\xi + tr\|} \leq \mu_r < \infty, \quad \xi + tr \in \Xi, \tag{28}$$

where $\|\cdot\|$ is the Euclidean norm and $t \in \mathbb{R}_+$. In particular, (28) follows for linear recourse programs with complete recourse, and this can always be ensured by using penalty formulations (Section 2.1.1). For unbounded support, the optimal distribution approximation is situated on the extreme points of Ξ, possibly extended by the extreme rays (see Fig. 6b). By the condition to preserve the first-order moment condition, we cannot extend along rays infinitely as long as the support is a polyhedral cone. The finite growth coefficients μ_r determine the upper bound on the objective value with respect to the ray extensions.

As an example, the unbounded space $\mathbb{R}^d$ can be described by conic polyhedral subsets in the form of every orthant (giving 2^d subsets), which are polyhedral cones. While there are large differences in the number of evaluation points based on the kind of support for the bounds discussed, there is also a complexity to consider with respect to how a high-dimensional space Ξ is partitioned. This

means there is a trade-off between the number of evaluation points for a subset and the number of partitioning subsets of the space.

3.3.6. Multistage case

The main challenge with the multistage extension of distribution approximation is that a discrete approximating distribution Q also defines a simplified information structure $\hat{\mathcal{F}}$ (a discrete tree) which interferes with the conditional evaluation of expectations. The fundamental question is whether we can apply the previously discussed bounds stage-wise: given a discretisation for the current stage, generate a discretisation for the next stage conditional on each discrete outcome. If possible, stage-wise discretisation significantly simplifies the discretisation, but this does not always suffice. The violation may occur when evaluating the conditional expectation of a stochastic quantity more than one stage ahead since this relies on the discretisation in intermediate stages that did not explicitly account for this. Under linear additive time-dependence with stage-wise independent noise, stage-wise discretisation is sufficient, while, in general, the entire process must be accounted for.

Distributional approximation on multistage stochastic programs has been explored by Edirisinghe and Ziemba (1992), Frauendorfer (1994, 1996), Frauendorfer et al. (2011) and Kuhn (2005). All of them use first-order and cross moment information and assume stage-wise independence or linear additive time dependence. More recently, Maggioni and Pflug (2019) have explored this for more general stochastic processes and show by counter-example that the entire process across all stages must then be accounted for during discretisation.

3.4. Function approximations

Function approximations are largely based on specific assumptions about the function at hand. The most general assumption is convexity which also holds for stochastic linear programs under right-hand side uncertainty, but some also deal with specific properties of these.

For function approximation to be useful, the expectation of the approximating function must also be easily evaluated. This is achieved by using simple function expressions (linear, piecewise linear or quadratic) and by ensuring separability. Separability means the approximating function over the high-dimensional random vector ξ can be expressed as a sum of functions of fewer (or single) elements of ξ. This reduces one high-dimensional integral to evaluating many low-dimensional integrals, which is a considerable simplification. Independence in the distribution would also ensure multiplicative separability.

For upper bounds, two kinds of function expressions have been considered to approximate $\phi(\xi)$ whose expectation is easy to evaluate. The aim is to fit these as tightly as possible within $\phi(\xi)$, whose procedures vary. First, there are separable piecewise linear functions

$$\tilde{\phi}(\xi) = u_0 + \sum_{i=1}^{d} u_i(\xi_i - \hat{\xi}_i)^+ + \sum_{i=1}^{d} u_{d+i}(\xi_i - \hat{\xi}_i)^-, \tag{29}$$

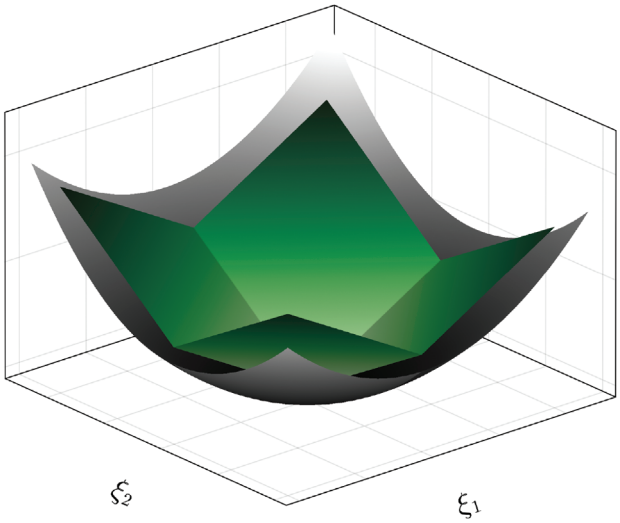

Fig. 7. A linear, separable and pointed cone, akin to a simple recourse function, that majorises a convex function.

where $\hat{\xi}$ is a predetermined value of ξ (the expectation, for instance) that defines the tip of the cone and $u \in \mathbb{R}^{2d+1}$ vector of coefficients. (See Fig. 7.) This approximating function primarily addresses growth conditions in the limits of ξ, and (29) may be refined by inserting a piecewise linear function within bounded intervals $[\xi_i^l, \xi_i^u]$ and letting the unbounded limit be determined by (29) outside of this interval (Birge and Wets, 1989). Second, there are second-order expressions

$$\tilde{\phi}(\xi) = u_0 + \sum_{i=1}^{d} u_i \xi_i + \sum_{i=1}^{d} u_{d+i} \xi_i^2, \tag{30}$$

whose primary aim is to alleviate the requirement of finding conditional expectations required by separable piecewise linear approximations. Consider, for example, that these are estimated from empirical data and that the confidence on second-order statistical estimates can be higher than for estimates of conditional expectations on many small subsets of the support. In fact, a requirement to find (30) is a (not necessarily separable) piecewise linear function which may be found using the procedures to find (29).

Lastly, there are penalty formulations of a two-stage stochastic linear program (SLP) suggested by Morton and Wood (1999), which give an analogous shape to (29), expressed as

$$\tilde{\phi}(\xi) = c_1(\omega)\hat{x}_1 + \mu^\top (B_1(\omega)x_0 + W_1(\omega)\hat{x}_1 - b_1(\omega))^+, \tag{31}$$

where $\hat{x}_1$ is a fixed decision inserted into the objective of a penalty formulation (Section 2.1), where $\hat{x}_1$ itself can be chosen such that it minimises the bound. This may not be separable to one-dimensional integrals (only lower dimensional) but can be applied more generally to (SLP) with random recourse. Observe that if only $b_1(\omega)$ is stochastic, (31) takes the same form as (29). Morton and Wood (1999) also showed that if we have dual formulations of the optimisation problem that

defines $\phi(\xi)$, we may apply similar bounds on the dual objective function to obtain lower bounds by weak duality. Such dual bounds are discussed in more detail in Section 4.3.

3.4.1. Piecewise linear approximating functions

The bound developed by Birge and Wallace (1988) (that generalises Wallace, 1987; Birge and Wets, 1989) assumes right-hand side uncertainty and bounded primal variables for two-stage (SLP) and finds a linear conic upper bounding approximation of the form (29) defined over unbounded support. (See Fig. 7.) The coefficients that determine the growth of this approximation are found by parametric evaluation of an altered formulation of the problem that requires a number of evaluations proportional to the dimensionality d of the support. It may not always be possible to find such finite slopes by the given procedure, in which case the upper bound evaluates to $+\infty$. Under penalty formulations of (SLP), the slopes are upper bounded by the penalty μ of constraint violation and guaranteed to give a finite upper bound. This also means this bound is tighter than (31) under the assumption of right-hand side uncertainty.

Powell and Frantzeskakis (1994) use a similar simplification as Wallace (1987) to get simplified upper bounding recourse functions of network problems (and discuss different strategies for doing so) but instead use these to optimise decisions. This allows for solving larger problems while still capturing the essence of the recourse problem at hand.

3.4.2. Second-order approximating functions

Dulá (1992) and Dulá and Murthy (1992) derived a second-order upper bound (30) that majorises a *known* piece-wise linear convex function ϕ. This known function is motivated by the piece-wise linear form of two-stage SLPs having either right-hand side or objective uncertainty. However, obtaining these explicitly requires extreme point enumeration of the (deterministic) dual or primal feasible sets. Other majorising approximations of this nature may also be used instead. Using *total* second-order information only (such that coefficients of second-order terms in (30) are equal), Dulá (1992) finds a closed form expression to determine the coefficients. This has no assumptions on dependence or boundedness of the support. Kall (1991) proposed an alternative perspective on this problem that leads to a non-smooth optimisation problem. Dulá and Murthy (1992) extend this to *marginal* second-order moments, which tightens the bound under the same assumptions. No closed-form solution is found, but the bound can be found by minimising a non-linear convex function in $d + 1$ variables with linear constraints.

Birge and Dulá (1991) use second-order information to bound a separable convex function ϕ, whose approximation $\tilde{\phi}$ is found by a line search. This directly extends the approach of Birge and Wallace (1988) but can also be applied to more general separable functions using its first-order derivatives.

3.5. Monte Carlo integration

The statistical counterpart to guaranteed integral bounds is to use Monte Carlo integration. That is, to sample an alternative distribution $\tilde{P}$ from the underlying distribution P to get a finite statistical estimator $\mathbb{E}^{\tilde{P}}[f(\tilde{x}, \xi)]$ of the true expectation $\mathbb{E}^{P}[f(\tilde{x}, \xi)]$ for a *fixed* candidate decision $\tilde{x}$.

Confidence intervals on this estimator can then derive statistical upper and lower bounds. This makes no assumption on the shape of $\xi \mapsto f(\tilde{x}, \xi)$, but it must be evaluated in a potentially very large number of sampled outcomes. The statistical properties rely on having a fixed decision, while optimisation with respect to $\mathbb{E}^{\tilde{P}}[f(x, \xi)]$ leads to bias in the estimator (see Section 4.2).

The average recourse value $\mathbb{E}^{\tilde{P}}[f(\tilde{x}, \xi)]$ is an estimator of the expected recourse value $\mathbb{E}^{P}[f(\tilde{x}, \xi)]$ which, by the central limit theorem, is asymptotically normally distributed, i.e.

$$\mathbb{E}^{\tilde{P}}[f(\tilde{x}, \xi)] \sim \mathrm{N}\left(\mathbb{E}^{P}[f(\tilde{x}, \xi)], \mathbb{V}\mathrm{ar}^{P}[f(\tilde{x}, \xi)]/\sqrt{|\tilde{P}|}\right), \tag{32}$$

where $|\tilde{P}|$ is the number of samples in $\tilde{P}$. Importantly, (32) holds for multistage problems where the samples are outcome paths (not necessarily trees) assuming the policy $\tilde{x}(\omega)$ provides decisions for every possible outcome path (Shapiro et al., 2014, Section 5.8.1). We may also extend by sampling the points of $\tilde{P}$ by more elaborate schemes to reduce variance in the estimator and get an improved convergence rate, referred to as quasi-Monte Carlo methods (Shapiro et al., 2014, Chapter 5.4).

4. Bounds on optimal expected values

We now turn to bounds on the optimal objective value $v(P)$. These are solely based on arguments from optimisation theory: primal-dual formulations, sub-optimality, relaxation and feasibility. Such bounds are particularly interesting in the context of determining if a given candidate decision is sufficiently close to the best obtainable decision. Given stochastic programs are fundamentally intractable to solve in their basic form, these are especially useful when pursuing approximation methods aimed at finding (close to) optimal decisions, and the bounds are used for validation.

This section discusses three approaches to bounding optimal objective values. In Section 4.1, we address candidate evaluation bounds based on sub-optimality, where the important concept is how to insert approximations into the original intractable formulation. In Section 4.2, we address how a relaxation of information structure can give simplified problems that provide bounds. Lastly, Section 4.3 reviews decision rules and aggregation bounds derived from these based on relaxation, tightening and sub-optimality. Using decision rules is also an alternative approximation approach to those based on discretising distributions to obtain scenario trees.

4.1. Bounds based on candidate policies

These bounds evaluate the objective function using candidate policies to get a bound on the optimal objective value $v(P)$. Namely, if $\tilde{x}(\omega)$ is a candidate primal policy and $\tilde{\lambda}(\omega)$ a candidate dual policy, we have that

$$\mathbb{E}^{P}[f(\tilde{x}(\omega), \xi(\omega))] \geq v(P) \geq v^{*}(P) \geq \mathbb{E}^{P}\big[f^{*}(\tilde{\lambda}(\omega), \xi(\omega))\big]. \tag{33}$$

International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies

Should $\tilde{x}(\omega)$ or $\tilde{\lambda}(\omega)$ be infeasible, we use the convention that this gives a primal or dual objective value of $\pm\infty$, which means these still give bounds but not very interesting ones. Infeasibility can be mitigated effectively by penalty formulations of stochastic programs (see Section 2.1.1) or by defining candidate policies according to rules that redeem feasibility. For the lower bound in (33) to be practical, the duality gap should be relatively tight and may require strong duality. For example, this is not the case for mixed-integer programs, but the upper bound still applies to these.

Obtaining candidate policies usually relies on some strategy of approximation. This is either done by discretisation into a scenario tree represented by a simplified information structure $\hat{\mathcal{F}}$ and alternative distribution Q, or by decision rules (Section 4.3). When using scenario trees, we must define an *extension rule* that determines the candidate policy for outcome paths ω not contained in the tree. These may consist of simpler or more elaborate rules but must adhere to the underlying information structure $\mathcal{F}$. Essentially, the extension rule in a given stage can only use information known at that stage. The concept of distances between trees is valuable in this context (Pflug and Pichler, 2012, 2014).

The simplest extension rule would be a nearest neighbour rule according to the distance from the tree to any given path $\tilde{\xi}_{[t]}$ up to its realisation at stage t where the policy must be determined. Regression procedures have been proposed by Keutchayan et al. (2017), while Stochastic Dual Dynamic Programming (SDDP) can be applied to obtain piece-wise linear policies for problems of a special kind (Pereira and Pinto, 1991). Alternatively, we may incrementally optimise the decision in each stage, conditional on decisions determined in past stages. This is mainly useful in the context of two-stage problems where the first-stage decision is fixed, and the second-stage problem is easily evaluated without the need for approximation of further stages. This is also the tightest possible candidate evaluation bound for two-stage problems where only the first-stage decision is sub-optimal. For multistage problems, however, the consecutive optimisation problems represent almost as difficult problems as the full problem, and we often resolve to simpler and less optimal extension rules, which means the quality of the extension procedure also interferes with the evaluation. Xu and Sen (2023) also introduce the concept of compromise policies that are determined by optimising a policy over an ensemble of value function approximations derived from many scenario trees (with extension rules) to exploit the principle of multiple replications instead of using larger scenario trees to obtain candidate policies.

The simplest scenario tree to construct a candidate policy from would be to use expected path, whose candidate evaluation bound is known as the *Expectation of the Expected Value* (EEV) solution (Madansky, 1960). Maggioni and Wallace (2012) and Crainic et al. (2018) also analyse how candidate decisions in two-stage problems can be obtained using partial information from the expected value solution while the remaining variables are optimised in the full problem (assuming a finite number of outcomes).

4.1.1. Out-of-sample evaluation bounds

Out-of-sample evaluation refers to the procedure of evaluating the performance of candidate policies found by some approximation procedure, like scenario generation, and then validating its performance in terms of the original objective by Monte Carlo integration (Section 3.5) to ease its evaluation of the original objective value. This means we replace P with a sample distribution $\tilde{P}$ assumed to be sufficiently large so that it represents P well. In the context of two-stage stochastic

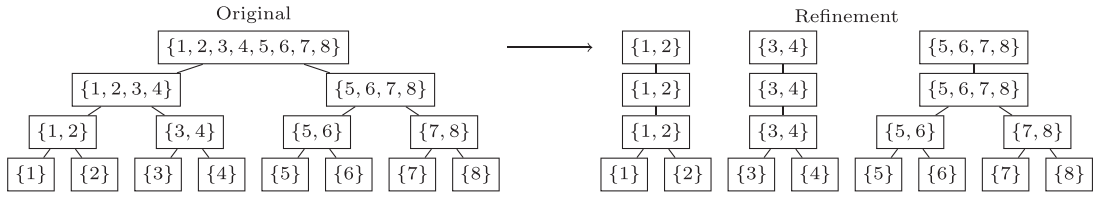

Fig. 8. Example of refinement of information structure.

programs, Kaut and Wallace (2007) proposed this as an approach to determine if scenario generation is effective by comparing the quality of candidate decisions obtained by different scenario generation methods. Out-of-sample evaluation essentially gauges different scenario generation procedures against each other by comparing which method can, by whichever means, produce the candidate decision that has the best evaluation bound. This has been a tool for the further development of scenario-generation methods. The concept of extension rules for out-of-sample evaluation in multistage problems proposed by Keutchayan et al. (2017) serves as a multistage counterpart. In principle, other evaluation bounds from Section 3 could also have been used to compare candidate decisions.

Evaluation of both the primal and dual objectives by candidate policies determined from scenario trees can gauge their representation of the underlying information structure in terms of the original objective function. While scenario set generation is mainly concerned with finding discrete outcomes, scenario trees additionally rely on effective representation of information structure. To our knowledge, this has not yet been explored in the context of scenario tree generation, but Kuhn (2008) used such primal-dual evaluation to gauge the effectiveness of decision rules (see Section 4.3). The advantage of primal-dual evaluation is that we get gap estimates of sub-optimality, providing a stopping criterion for sufficient approximation quality. In contrast, comparing upper bounds solely determines if one approximation approach is better than the other. Effective discrete representation of information structure tuned to the problem at hand is an interesting avenue of future research that currently has received little attention in the research literature on scenario tree generation.

4.2. Bounds on refinement of information structure

Refinement of information structure implies splitting a tree into smaller sub-trees and optimising decisions on these instead. This provides a collection of simpler optimisation problems whose solutions combine to provide a lower bound on the original problem. A conceptual illustration is given in Fig. 8. This section presents a general setup that applies to various bounds in the literature derived from partitioning, group sub-problems and sampling. However, note that some of these approaches involve combinatorial aspects not explained in detail. We also do not rely on the value of ξ to work out these bounds and, hence, refer to ω instead.

Throughout this section, we assume a finite discrete distribution $\tilde{P}$ with support $\tilde{\Omega}$ such that the optimal objective value and solution set of $\mathbb{E}^{\tilde{P}}[f(x(\omega), \xi(\omega))]$ is sufficiently close to those of $\mathbb{E}^{P}[f(x(\omega), \xi(\omega))]$ (Shapiro et al., 2014, Section 7.2.5). We may, for example, interpret $\tilde{P}$ as a

sufficiently large sample from P. Sampling (in a stage-wise manner) or using some other discretisation procedure to obtain $\tilde{P}$ is also the approach used in practice for computing these bounds.

The refinement is constructed by finding a (countable) collection of alternative distributions $\{Q_k\}_{k\in\mathcal{K}}$ on subsets $\{\Omega_k\}_{k\in\mathcal{K}}$ of $\tilde{\Omega}$ and a convex combination of positive weights $\{\gamma_k\}_{k\in\mathcal{K}}$ (where $\sum_{k\in\mathcal{K}}\gamma_k = 1$) such that

$$\tilde{P}(\omega) = \sum_{k\in\mathcal{K}} \gamma_k Q_k(\omega), \tag{34}$$

where the alternative distributions compose the original distribution. Supports Ω_k that are strictly smaller than $\tilde{\Omega}$ is most interesting since this implies problems $\mathbb{E}^{Q_k}[f(x,\xi)]$ are easier to solve than the original formulation using $\tilde{P}$. Probability assignment in terms of $\{Q_k\}_{k\in\mathcal{K}}$ and subset selection in terms of $\{\Omega_k\}_{k\in\mathcal{K}}$ are two alternative interpretations of how the distribution $\tilde{P}$ is simplified; however, note that the relative assignment of probabilities within Ω_k is solely determined by Q_k. By linearity of expectations in their probability assignment, we have that

$$\mathbb{E}^{\tilde{P}}[f(x(\omega),\xi(\omega))] = \sum_{k\in\mathcal{K}} \gamma_k \mathbb{E}^{Q_k}[f(x(\omega),\xi(\omega))], \tag{35}$$

where the latter expression represents a convex combination of sub-problem evaluations. By minimising (35) with respect to decision x, we obtain the lower bound

$$\min_{x(\omega)\in\mathcal{X}(\omega)} \left\{\mathbb{E}^{\tilde{P}}[f(x(\omega),\xi(\omega))]\right\} \geq \sum_{k\in\mathcal{K}} \gamma_k \min_{x(\omega)\in\mathcal{X}(\omega)} \left\{\mathbb{E}^{Q_k}[f(x(\omega),\xi(\omega))]\right\} \tag{36}$$

by interchanging the order of optimisation and summation over $\mathcal{K}$. This also has the interpretation that the optimal value map $\tilde{P} \mapsto v(\tilde{P})$ is concave in $\tilde{P}$ (Maggioni and Pflug, 2016, Lemma 2.1). This is a simplification of the original formulation since we may solve smaller batches $k \in \mathcal{K}$ of sub-problems instead of the full problem.

A consequence of solving the stochastic program in batches of sub-problems $\mathbb{E}^{Q_k}[f(x(\omega),\xi(\omega))]$ is that this *relaxes the underlying information structure*, and leads to batch-specific optimal decisions that cannot be implemented in the original formulation. The fact that this leads to a lower bound can be illustrated conceptually (Fig. 9) by considering how the evaluation of a convex combination of batch-specific objective functions $\mathbb{E}^{Q_k}[f(x(\omega),\xi(\omega))]$ compares to an evaluation of each of them.

Bounds based on partitioning. The first (and classic) example of refinement bounds is the expectation over all wait-and-see solutions (decisions are determined *after* uncertainty is revealed) by Madansky (1960). This follows from interchanging the order of minimisation and integration

$$\min_{x(\omega)\in\mathcal{X}(\omega)} \left\{\mathbb{E}^{\tilde{P}}[f(x(\omega),\xi(\omega))]\right\} \geq \mathbb{E}^{\tilde{P}}\left[\min_{x(\omega)\in\mathcal{X}(\omega)} \{f(x(\omega),\xi(\omega))\}\right], \tag{37}$$

where the sub-groups are considered to be the atoms of $\tilde{\Omega}$. Partitioning into larger subsets of $\tilde{\Omega}$ than the atoms strictly strengthens this bound (Maggioni and Pflug, 2016, Section 2). We

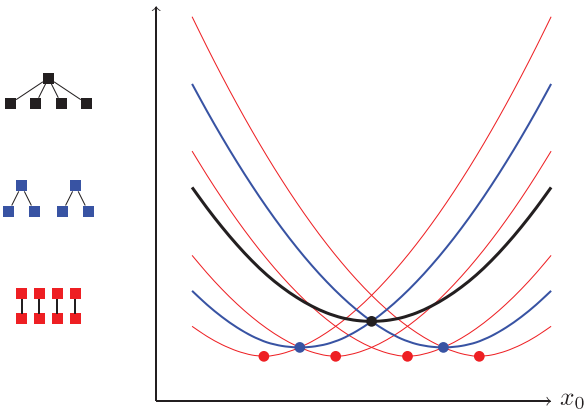

Fig. 9. Illustration of different objective functions $\mathbb{E}^{Q_k}[f(x(\omega), \xi(\omega))]$ in terms of first-stage decision x_0 under refinement of information structure. Re-enforcing non-anticipativity in the first-stage decision corresponds to a convex combination of the refined objective functions. Observe that each objective function's minimiser (dots) changes according to the information structure.

may also make a refinement chain of consecutive coarser partitions, where larger alternative distributions Q_k are convex combinations of smaller ones, to get a hierarchy of bounds that get progressively stronger.

Bounds based on group sub-problems. We may instead define sub-groups of Ω that can also intersect each other. The associated alternative distributions Q_k must still adhere to (34). Birge (1982) derived the first such result referred to as the sum of pairs sub-problems where a particular reference outcome ω_r is paired with every other outcome such that $\Omega_k = \{\omega_k, \omega_r\}$ for all $k \neq r$. The distributions $Q_k(\omega_r)$ are adjusted such that the overall probability is consistently given as $\tilde{P}(\omega_r) = \sum_k Q_k(\omega_r)$. The reference outcome ω_r here represents a particularly important outcome that the decision-maker always wants to account for. This grouping gives a lower bound on $v(P)$ but is also an upper bound on the wait-and-see solution (37). The generalisation instead chooses $\{\Omega_k^n\}_{k \in \mathcal{K}}$ to be all subsets of $\tilde{\Omega}$ of a given size n, referred to as the *group sub-problems*. These give a lower bound for each size n that is monotonously non-decreasing in n, with edge cases that coincide with the wait-and-see problems and the original problem. The formulation such that the original problem is an edge case was derived by Sandıkçı et al. (2013) and Maggioni et al. (2014, 2016) for two-stage and multistage problems.

Sampling-based bounds. Monte Carlo sampling is commonly used as a viable option for solving intractable stochastic programs when the distribution is too large. Mak et al. (1999) showed that a sampled distribution Q_k^n of a given size n (which may represent historical data itself) provides a lower-bounding optimal objective value $v(Q_k^n)$ *in expectation* over many samples of Q_k^n. This implies that solving a stochastic program on historical data gives, on average, a lower bounding objective value with respect to the underlying distribution. Furthermore, Mak et al. (1999, Theorem 2) show

the lower bound is monotonously increasing with the sample size n. In this setting, each alternative distribution Q_k^n represents a sampled empirical distribution of a given size, while the weighting γ_k represents the overall probability of obtaining this empirical distribution in particular. Over many sampled distributions Q_k^n, their expected relative occurrence is γ_k. Sample average approximation is analogous to group sub-problems where the relative weights $\{\gamma_k\}_{k\in\mathcal{K}}$ have an additional sampling error, and potentially also that outcomes are defined to be equiprobable in each Q_k^n according to empirical distributions. There is much more to be said about solving stochastic programs with respect to sampled distributions, for which we refer to Bayraksan and Morton (2009) and Shapiro et al. (2014).

4.3. Bounds based on decision rules and aggregation

In their basic form, stochastic programs are infinite mathematical programs. Aggregation of variables and constraints was first introduced in finite (deterministic) linear programming as a simplification to make them smaller and easier to solve. In stochastic programming, however, aggregation turns an infinite linear program into a finite one by integrating (aggregating) the stochastic argument out of the problem, which is a considerable improvement (to say the least). In this context, we first choose a decision rule defined as a linear combination of basic rules that are predefined functions of the stochastic data process ξ (or ω). Inserting the decision rule in place of a general policy in the stochastic program allows evaluating expectations first and instead optimising over the finite number of coefficients in the linear combination. In contrast, a general policy must be optimised for all (infinite number of) outcomes.

Theory on aggregation bounds (that stem from deterministic linear programming) allows deriving bounds on the original infinite formulation of the stochastic program in terms of the solvable finite formulation. These bounds use arguments of how decision rules lead to relaxation, tightening or sub-optimality with respect to the original formulation. Primal-dual pairs give rise to upper and lower bounds according to weak duality. Constraint aggregation in the primal corresponds to variable aggregation in the dual, and conversely. Aggregation may also be done over stages in multistage formulations.

Aggregation bounds applied to stochastic programming started with Zipkin (1980a, 1980b) who analysed aggregation of variables and constraints for finite (deterministic) linear programs. Birge (1985) extended these results in the setting of stochastic programming by instead aggregating over uncertainty or stages assuming right-hand side uncertainty. Wright (1994) generalised this in a measure-theoretic framework using Lagrangian duality to general stochastic linear programs in the context of coarsened (simplified) information structures. More generally, this can be formulated by decision rules for which similar results apply. Decision rules have been successfully applied in the context of robust optimisation and were, for this reason, re-introduced to stochastic programming by Shapiro and Nemirovski (2005) as a means of tractable complexity reduction. Kuhn et al. (2011) have shown how to find solutions to stochastic linear programs where either the primal or the dual policy is replaced by decision rules (leading to semi-infinite programs), and aggregation bounds provide the optimality gap of these decision rules with respect to the original formulation.

4.3.1. Statement of decision rules

We state the formulation of decision rules and, for simplicity of exposition, we do this in the context of stochastic linear programs (SLP). Decision rules can be applied to replace the primal policy $x(\omega)$, the dual policy $\lambda(\omega)$, or both. Using decision rules in the dual policy also has the alternative interpretation of constraint aggregation in the primal problem, and conversely.

Consider that we define decision rules $\hat{x}(\omega)$ for primal decisions and $\hat{\lambda}(\omega)$ for dual decisions. These are composed of linear combinations of simpler fixed policies

$$\hat{x}(\omega) = \sum_{k=1}^{K} w_k^x h_k^x(\omega), \qquad \hat{\lambda}(\omega) = \sum_{l=1}^{L} w_l^\lambda h_l^\lambda(\omega), \tag{38}$$

where $h^x(\omega) := \{h_k^x(\omega)\}_{k=1,\dots,K}$ and $h^\lambda(\omega) := \{h_l^\lambda(\omega)\}_{l=1,\dots,L}$ are predefined functions of ω, while w^x and w^λ are vectors of K and L elements, respectively, that define the linear combinations. The functions $h^x(\omega)$ and $h^\lambda(\omega)$ must adhere to the same information structure $\mathcal{F}$ and have the same dimensionality as the policies x and λ in the original formulation. Simplification from general policy to a decision rule effectively *limits its flexibility* since it is reduced to a linear combination of predefined functions. Furthermore, as noted by Kuhn (2008), inserting a decision rule for constraint multipliers corresponds to *more flexibility* in decisions.

To see how decision rules cause ω to be aggregated out of the problem, consider the Lagrangian of (SLP)

$$L(x(\omega), \lambda(\omega)) = \mathbb{E}^P\big[c(\omega)^\top x(\omega) - \lambda(\omega)^\top (A(\omega)x(\omega) - b(\omega))\big], \tag{39}$$

where inserting the decision rules $\hat{x}(\omega)$, $\hat{\lambda}(\omega)$ in place of the general policies $x(\omega)$, $\lambda(\omega)$ gives

$$L(\hat{x}(\omega), \hat{\lambda}(\omega)) = \hat{c}^\top w^x - (w^\lambda)^\top (\hat{A} w^x - \hat{b}), \tag{40}$$

where

$$\hat{c}_k = \mathbb{E}^P\big[c(\omega)^\top h_k^x(\omega)\big], \qquad \hat{A}_{lk} = \mathbb{E}^P\big[h_l^\lambda(\omega)^\top A(\omega) h_k^x(\omega)\big], \qquad \hat{b}_l = \mathbb{E}^P\big[h_l^\lambda(\omega)^\top b(\omega)\big],$$

are the aggregated parameters. In the usual sense of linear programming, there are then finite primal and dual LPs associated with (40). Since the reliance on ω is limited to predefined functions, evaluation of expectations can be performed immediately and aggregates ω out of the problem. We may then proceed to optimise the decision rule with respect to the aggregated parameters and linear coefficients instead. Using decision rules either for variables or for constraint multipliers leads to aggregation bounds on the optimal value $v(P)$ of the original formulation (see Section 4.3.3), which we illustrate in Section 4.3.2.

There is also a strong link between decision rules and function approximation (Section 3.4), where the only conceptual difference is a (sometimes trivial) evaluation of the objective function in terms of the decision rule. Consequently, decision rules and aggregation bounds have close analogies to those in Section 3.4.

4.3.2. Example: Variable and constraint aggregation

Variable aggregation can be interpreted as replacing many variables with a single one. Consider the primal constraint set of (SLP),

$$A(\omega)x(\omega) \geq b(\omega), \quad \forall \omega \in \Omega,$$

where, for a single outcome, there is only a single constraint that the general policy x must fulfil in that outcome. If we instead use a decision rule $\hat{x}(\omega) = w^x h^x(\omega) \geq 0$ and insert it into the constraint, we have that

$$w^x A(\omega) h^x(\omega) \geq b(\omega), \quad \forall \omega \in \Omega,$$

and the single coefficient w^x is subject to constraints for *all* $\omega \in \Omega$. Hence, the decision rule led to tightened constraints compared to the general policy because the flexibility to adjust to each outcome ω has been limited. The decision rule also relies on the function $h^x(\omega)$ that was chosen to decide how these constraints can be fulfilled.

Constraint aggregation can be interpreted as making a (weighted) sum of constraints. If we have two valid inequality constraints and add them together (possibly by positive weights), we obtain a new inequality constraint that must also be valid. If we only know that the summed constraint is valid, however, we cannot say anything about the validity of each constraint that composes it. This means the summation of constraints results in relaxation. To see why constraints are summed, consider that we use a decision rule $\hat{\lambda}(\omega) = w^\lambda h^\lambda(\omega) \geq 0$ for the constraint multipliers in SLP. The restatement of SLP under Lagrangian relaxation of its constraints is

$$\min_{x(\omega)\geq 0} \mathbb{E}^P\big[c(\omega)^\top x(\omega) - \lambda(\omega)^\top (A(\omega)x(\omega)) - b(\omega)\big],$$

and inserting the decision rule $\hat{\lambda}(\omega)$ instead gives

$$\min_{x(\omega)\geq 0} \mathbb{E}^P\big[c(\omega)^\top x(\omega)\big] - w^\lambda \mathbb{E}^P\big[h^\lambda(\omega)^\top (A(\omega)x(\omega) - b(\omega))\big],$$

whose corresponding non-relaxed constraints are

$$\mathbb{E}^P\big[h^\lambda(\omega)^\top A(\omega)x(\omega)\big] \geq \mathbb{E}^P\big[h^\lambda(\omega)^\top b(\omega)\big],$$

with dual multiplier w^λ, and where the function $h^\lambda(\omega)$ acts as weighting in the summation (integration) of constraints. This also means we replace many constraint multipliers with a single one, in analogy to variable aggregation. The relaxation follows from the fact that only the aggregated constraint must be fulfilled by $x(\omega)$, which gives it more flexibility.

4.3.3. Aggregation bounds

We now consider using decision rules to replace either primal or dual policies to obtain bounds on the optimal objective value. Let $\mathbb{P}(x, \lambda)$ denote the primal problem (SLP) with respect to the variables in the first argument and constraint multipliers in the second argument. General policies are denoted by $x(\omega)$ and $\lambda(\omega)$ while decision rules (38) inserted in their place are denoted by $\hat{x}(\omega)$

and $\hat{\lambda}(\omega)$. For ease of readability, we omit their argument ω in this section. Variable aggregation in the primal corresponds to constraint aggregation in the dual (as well as the converse), and so $\mathbb{D}(\lambda, x)$ denotes the dual problem of $\mathbb{P}(x, \lambda)$ with respect to the same arguments. Furthermore,

$$\inf_{x \in \mathcal{X}} \mathbb{P}(x, \lambda) \geq \sup_{\lambda \in \Lambda} \mathbb{D}(\lambda, x), \tag{41}$$

by weak duality. Let $\mathcal{X} = \text{feas}(\mathbb{P}(x, \lambda))$ and $\Lambda = \text{feas}(\mathbb{D}(\lambda, x))$ denote the feasible region of decisions in $\mathbb{P}(x, \lambda)$ and $\mathbb{D}(\lambda, x)$, respectively. Variable aggregation tightens the feasible region, while constraint aggregation relaxes it. We then have the relations,

$$\text{feas}(\mathbb{P}(\hat{x}, \lambda)) \subseteq \begin{matrix} \text{feas}(\mathbb{P}(x, \lambda)) \\ \text{feas}(\mathbb{P}(\hat{x}, \hat{\lambda})) \end{matrix} \subseteq \text{feas}(\mathbb{P}(x, \hat{\lambda})), \tag{42}$$

and

$$\text{feas}(\mathbb{D}(\lambda, \hat{x})) \supseteq \begin{matrix} \text{feas}(\mathbb{D}(\lambda, x)) \\ \text{feas}(\mathbb{D}(\hat{\lambda}, \hat{x})) \end{matrix} \supseteq \text{feas}(\mathbb{D}(\hat{\lambda}, x)). \tag{43}$$

Observe that inserting one kind of decision rule at a time gives simple relations between feasible regions, but there is no simple relation between the original and the fully simplified problems. We have the following relations between objective values

$$\mathbb{P}(\hat{x}, \lambda) \geq \inf \mathbb{P}(\hat{x}, \lambda) \geq \begin{matrix} \inf \mathbb{P}(x, \lambda) \geq \sup \mathbb{D}(\lambda, x) \\ \inf \mathbb{P}(\hat{x}, \hat{\lambda}) \geq \sup \mathbb{D}(\hat{\lambda}, \hat{x}) \end{matrix} \geq \sup \mathbb{D}(\hat{\lambda}, x) \geq \mathbb{D}(\hat{\lambda}, x), \tag{44}$$

where $\mathbb{P}(\hat{x}, \lambda)$ and $\mathbb{D}(\hat{\lambda}, x)$ denote objective evaluation by a feasible decision policy.

The underlying formulations $\mathbb{P}(x, \lambda)$ and $\mathbb{D}(\lambda, x)$ are considered to be unsolvable, but bounds on these can be derived using decision rules according to the inequalities (44). Consider also that a low gap between these bounds means the predefined form of decision rules (38) is effective for the original formulation. Indeed, there exist more or less appropriate functions $h^x(\omega)$ and $h^\lambda(\omega)$ to obtain good results with low optimality gaps.

Tractable reformulations of $\inf \mathbb{P}(\hat{x}, \lambda)$ and $\sup \mathbb{D}(\hat{\lambda}, x)$ into linear and semi-definite programs of moderate sizes that can be solved directly have been laid out by Kuhn et al. (2011). Alternatively, we may solve the finite formulations $\mathbb{P}(\hat{x}, \hat{\lambda})$ and $\mathbb{D}(\hat{\lambda}, \hat{x})$ to obtain candidate primal $\tilde{x}(\omega)$ and dual policies $\tilde{\lambda}(\omega)$ and evaluate these within the objectives of $\mathbb{P}(\hat{x}, \lambda)$ and $\mathbb{D}(\hat{\lambda}, x)$ to get bounds on $v(P)$. This procedure is also described in Section 4.1 regarding scenario trees and extension rules.

5. Discussion and conclusions

Stochastic programming is applied to make effective decisions for real-world problems under uncertainty. While there is something fundamentally intractable in their formulations, bounds can give valuable insights into how well a given solution performs within the original formulation, or bounding distributions can be used as approximations in themselves.

The insight that function approximation and distribution approximation are tightly connected is a profound observation. This stems from the duality between functions and probability distributions, a connection seen throughout the paper. These perspectives were used for different approaches to evaluation bounds in Section 3, but also have close analogies to different ways of finding approximate solutions to stochastic programs either by scenario trees (distributions) or decision rules (functions) in Section 4.

The fundamental insight of evaluation bounds in Section 3 is that finite approximate distributions can be used to obtain bounds, even if the underlying distribution is continuous. This is a consequence of the GMP, solved by discrete distributions whose moments coincide with the underlying distribution. Furthermore, by the assumption of convexity in terms of the stochastic argument ξ, we may find analytical expressions for discrete distributions that provide guaranteed bounds. Bounds on optimal objective values in Section 4 also complement the evaluation bounds in Section 3. This is valuable both for applying and developing methods to solve such problems.

Discussion: bounds and modelling

Counter-intuitive mistakes may occur when one aims to make an approximation of uncertainty while actually getting bounds. These insights are especially relevant when applying models to decision problems under uncertainty, even if one does not solve a stochastic program.

An intuitive way to make a deterministic approximation to uncertainty is to insert the expected value of uncertain parameters. While the expectation of the stochastic phenomenon (i.e., ξ) can be used as a best estimate for the phenomenon itself, stochastic programming deals with a non-linear transformation of the stochastic variables through the objective function, and what we really aim to find is the *expected objective value*. Hence, caution is due. For objective functions that are convex in the stochastic parameters, using the expected outcome of the phenomenon instead gives a lower bound (Section 3.3). Not only will this give a bound, but it will also be optimistic, the opposite of what we might want. More worryingly, an objective function that is convex-concave in uncertain parameters may lead to conservatism with respect to some parameters and optimism with respect to others, implicitly making a priority.

Consider a two-stage production planning problem as an illustration. The objective of the second-stage program may then be convex in its uncertain capacity and concave in the uncertain cost of materials, while additional operational considerations also limit flexibility in production. Using only the expected value means we overestimate the utility of additional capacity, given that operational considerations may cause other bottlenecks. We also underestimate how flexibility in production can counteract increased material costs by scaling down the production of products that turn out unprofitable. This means we are optimistic with respect to capacity and pessimistic with respect to the cost of materials.

Another intuitive way of approximating uncertainty is to plan for a restricted small set of outcomes instead of (approximately) every outcome, like sampling from the distribution. As explored in Section 4.2, considering only a restricted subset of outcomes at a time (but overall considering all of them) will, in combination, lead to a lower bound on the optimal objective value. Again, this is optimistic. One interpretation of this is that the optimal decision overestimates its availability of information and is less prepared for outcomes that were not explicitly accounted for. The

recommended counter-measure in such situations is always to do candidate evaluation (Section 4.1) after a solution has been found to get a conservative estimate of how well a given decision or policy really performs in terms of the original formulation.

Scenario analysis (analysing decisions for one outcome at a time) is another case of apparent approximation of uncertainty that does not really build consideration of uncertainty into decisions. The issue is that considering a single outcome, we get the impression of knowing the future perfectly when it is actually unknowable (by definition of it being stochastic). Considering a single outcome at a time will *not* lead to decisions that buy options and enable flexibility to deal with multiple potential outcomes of the future and are thus qualitatively different (Wallace, 2010). The gap to the wait-and-see lower bound in Section 4.2 quantifies the error that results from scenario analysis.

Future research

A significant challenge for stochastic programming is how to solve multistage problems in a tractable manner. Using bounds to validate solutions can be valuable for further development in this domain. Approximations in the form of scenario trees or decision rules can potentially provide reasonable solutions by parsimonious representations but do not necessarily give indications of optimality or correct estimates of its optimal objective value. This is why bounds can help support their development.

Evaluation bounds (Section 3) have received much attention in the past while, more recently, the literature has shifted towards approximations in the form of scenario generation. Both are concerned with discretising distributions, but scenario generation has an added layer of complication since it primarily concerns the quality of solutions obtained from *solving* a stochastic program with a given discrete distribution. The property of convex stochastic dominance (Section 3.2) conserves bounds for all decisions, which is an interesting link between these. Still, we believe there are synergies to be explored between bounds and scenario generation that can provide valuable insights. The concept of effective and ineffective scenarios developed in the setting of distributionally robust optimisation (Rahimian et al., 2019, 2022) may also provide further such connections.

There has been much recent research on risk aversion and aversion to ambiguity in the representation of uncertainty itself (not knowing the distribution). These are valuable tools to build conservatism into decisions and are tightly linked to bounds in the sense that they are often formulated to minimise an upper bound on the objective evaluation, subject to ambiguity (more general than moment conditions). These problems are challenging to solve (even more so for multistage problems), and quantifying interval estimates to their optimal solution is valuable to improve these approaches further and to give valuable insights for real applications. The literature is sparse in this domain but growing. Future work could also include using bounds from Section 4 within solution algorithms.

References

Bayraksan, G., Morton, D.P., 2009. Assessing solution quality in stochastic programs via sampling. In *Decision Technologies* and Applications. INFORMS, INFORMS TutORials in Operations Research, chapter 5, pp. 102–122.

Ben-Tal, A., Hochman, E., 1972. More Bounds on the expectation of a convex function of a random variable. *Journal of Applied Probability* 9, 4, 803–812. 3212616.

Bertsimas, D., Tsitsiklis, J., 1997. *Introduction to Linear Programming*. Athena Scientific, Nashua, NH.

Birge, J.R., 1982. The value of the stochastic solution in stochastic linear programs with fixed recourse. *Mathematical Programming* 24, 1, 314–325.
Birge, J.R., 1985. Aggregation bounds in stochastic linear programming. *Mathematical Programming* 31, 1, 25–41.
Birge, J.R., Dulá, J.H., 1991. Bounding separable recourse functions with limited distribution information. *Annals of Operations Research* 30, 1, 277–298.
Birge, J.R., Louveaux, F., 2011. *Introduction to Stochastic Programming*. (2nd edn). Springer Series in Operations Research and Financial Engineering. Springer New York, New York.
Birge, J.R., Wallace, S.W., 1988. A separable piecewise linear upper bound for stochastic linear programs. *SIAM Journal on Control and Optimization* 26, 3, 725–739.
Birge, J.R., Wets, R.J.B., 1986. Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse. In Prékopa, A. and Wets, R.J.B., (eds), *Stochastic Programming 84 Part I. Mathematical Programming Studies*, Springer, Berlin, pp. 54–102.
Birge, J.R., Wets, R.J.B., 1987. Computing bounds for stochastic programming problems by means of a generalized moment problem. *Mathematics of Operations Research* 12, 1, 149–162.
Birge, J.R., Wets, R.J.B., 1989. Sublinear upper bounds for stochastic programs with recourse. *Mathematical Programming* 43, 1, 131–149.
Crainic, T.G., Maggioni, F., Perboli, G., Rei, W., 2018. Reduced cost-based variable fixing in two-stage stochastic programming. *Annals of Operations Research*. http://10.1007/s10479-018-2942-8.
Dokov, S.P., Morton, D.P., 2002. higher-order upper bounds on the expectation of a convex function. Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II, Institut für Mathematik.
Dokov, S.P., Morton, D.P., 2005. Second-order lower bounds on the expectation of a convex function. *Mathematics of Operations Research* 30, 3, 662–677.
Dulá, J.H., 1992. An upper bound on the expectation of simplicial functions of multivariate random variables. *Mathematical Programming* 55, 1, 69–80.
Dulá, J.H., Murthy, R.V., 1992. A Tchebysheff-type bound on the expectation of sublinear polyhedral functions. *Operations Research* 40, 5, 914–922.
Dupačová, J., 1966. On minimax solutions of stochastic linear programming problems. *Časopis pro pěstování matematiky* 91, 4, 423–430.
Dupačová, J., 1987. The minimax approach to stochastic programming and an illustrative application. *Stochastics* 20, 1, 73–88.
Edirisinghe, N.C.P., 1996. New second-order bounds on the expectation of saddle functions with applications to stochastic linear programming. *Operations Research* 44, 6, 909–922.
Edirisinghe, N.C.P., Ziemba, W.T., 1992. Tight bounds for stochastic convex programs. *Operations Research* 40, 4, 660–677.
Edirisinghe, N.C.P., Ziemba, W.T., 1994a. Bounding the expectation of a saddle function with application to stochastic programming. *Mathematics of Operations Research* 19, 2, 314–340.
Edirisinghe, N.C.P., Ziemba, W.T., 1994b. Bounds for two-stage stochastic programs with fixed recourse. *Mathematics of Operations Research* 19, 2, 292–313. 3690222.
Edirisinghe, N.C.P., Ziemba, W.T., 1996. Implementing bounds-based approximations in convex-concave two-stage stochastic programming. *Mathematical Programming* 75, 2, 295–325.
Edmundson, H.P., 1957. Bounds on the expectation of a convex function of a random variable. Technical report, RAND Corporation, Santa Monica, CA.
Frauendorfer, K., 1988. Solving SLP recourse problems with arbitrary multivariate distributions—the dependent case. *Mathematics of Operations Research* 13, 3, 377–394.
Frauendorfer, K., 1989. A simplicial approximation scheme for convex two-stage stochastic programming problems. *Manuscript, IOR University of Zürich* 18, 177–205.
Frauendorfer, K., 1992. Barycentric Approximation. In Frauendorfer, K. (ed.), *Stochastic Two-Stage Programming. Lecture Notes in Economics and Mathematical Systems*, Springer, Berlin, pp. 67–123.
Frauendorfer, K., 1994. Multistage stochastic programming: error analysis for the convex case. *Zeitschrift für Operations Research* 39, 1, 93–122.
Frauendorfer, K., 1996. Barycentric scenario trees in convex multistage stochastic programming. *Mathematical Programming* 75, 2, 277–293.

Frauendorfer, K., Kuhn, D., Schürle, M., 2011. Barycentric bounds in stochastic programming: theory and application. In Infanger, G. (ed.), *Stochastic Programming: The State of the Art In Honor of George B. Dantzig. International Series in Operations Research & Management Science*, Springer, New York, pp. 67–96.

Gassmann, H., Ziemba, W.T., 1986. A tight upper bound for the expectation of a convex function of a multivariate random variable. In Prékopa, A. and Wets, R.J.B. (eds), *Stochastic Programming 84 Part I. Mathematical Programming Studies*, Springer, Berlin, pp. 39–53.

Huang, C.C., Ziemba, W.T., Ben-Tal, A., 1977. Bounds on the expectation of a convex function of a random variable: with applications to stochastic programming. *Operations Research* 25, 2, 315–325.

Jensen, J.L.W.V., 1906. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica* 30, 175–193.

Kall, P., 1987. Stochastic programs with recourse: an upper bound and the related moment problem. *Zeitschrift für Operations Research* 31, 3, A119–A141.

Kall, P., 1988. Stochastic programming with recourse: Upper bounds and moment problems - a review. In *Advances in Mathematical Optimization. Mathematical Research*, Vol. 45, De Gruyter, Berlin, pp. 86–103.

Kall, P., 1991. An upper bound for SLP using first and total second moments. *Annals of Operations Research* 30, 1, 267–276.

Kall, P., Wallace, S.W., 1994. *Stochastic Programming*. John Wiley & Sons, Chichester.

Karr, A.F., 1983. Extreme points of certain sets of probability measures, with applications. *Mathematics of Operations Research* 8, 1, 74–85.

Kaut, M., Wallace, S.W., 2007. Evaluation of scenario-generation methods for stochastic programming. *Pacific Journal of Optimization* 3, 2, 257–271.

Kemperman, J.H.B., 1968. The general moment problem, a geometric approach. *The Annals of Mathematical Statistics* 39, 1, 93–122.

Keutchayan, J., Gendreau, M., Saucier, A., 2017. Quality evaluation of scenario-tree generation methods for solving stochastic programming problems. *Computational Management Science* 14, 3, 333–365.

King, A.J., Wallace, S.W., 2012. Modeling with Stochastic Programming. *Springer Series in Operations Research and Financial Engineering*. Springer, New York, NY.

Klein Haneveld, W.K., 1986. Some linear programs in probabilities and their duals. In Klein Haneveld, W.K. (ed.), *Duality in Stochastic Linear and Dynamic Programming. Lecture Notes in Economics and Mathematical Systems*, Springer, Berlin, pp. 49–111.

Kuhn, D., 2005. Barycentric approximation scheme. In Kuhn, D., Beckmann, M., Künzi, H.P., Fandel, G., Trockel, W., Basile, A., Drexl, A., Dawid, H., Inderfurth, K., Kürsten, W. and Schittko, U. (eds), *Generalized Bounds for Convex Multistage Stochastic Programs*. Lecture Notes in Economics and Mathematical Systems, Springer, Berlin, pp. 51–81.

Kuhn, D., 2008. Aggregation and discretization in multistage stochastic programming. *Mathematical Programming* 113, 1, 61–94.

Kuhn, D., Wiesemann, W., Georghiou, A., 2011. Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming* 130, 1, 177–209.

Madansky, A., 1959. Bounds on the expectation of a convex function of a multivariate random variable. *The Annals of Mathematical Statistics* 30, 3, 743–746.

Madansky, A., 1960. Inequalities for stochastic linear programming problems. *Management Science* 6, 2, 197–204.

Maggioni, F., Allevi, E., Bertocchi, M., 2014. Bounds in multistage linear stochastic programming. *Journal of Optimization Theory and Applications* 163, 1, 200–229.

Maggioni, F., Allevi, E., Bertocchi, M., 2016. Monotonic bounds in multistage mixed-integer stochastic programming. *Computational Management Science* 13, 3, 423–457.

Maggioni, F., Pflug, G.C., 2016. Bounds and approximations for multistage stochastic programs. *SIAM Journal on Optimization* 26, 1, 831–855.

Maggioni, F., Pflug, G.C., 2019. Guaranteed bounds for general nondiscrete multistage risk-averse stochastic optimization programs. *SIAM Journal on Optimization* 29, 1, 454–483.

Maggioni, F., Wallace, S.W., 2012. Analyzing the quality of the expected value solution in stochastic programming. *Annals of Operations Research* 200, 1, 37–54.

Mak, W.K., Morton, D.P., Wood, R., 1999. Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters* 24, 1-2, 47–56.

Morton, D.P., Wood, R.K., 1999. Restricted-recourse bounds for stochastic linear programming. *Operations Research* 47, 6, 943–956.

Pereira, M.V.F., Pinto, L.M.V.G., 1991. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming* 52, 1-3, 359–375.

Pflug, G.C., Pichler, A., 2012. A distance for multistage stochastic optimization models. *SIAM Journal on Optimization* 22, 1, 1–23.

Pflug, G.C., Pichler, A., 2014. Multistage Stochastic Optimization. *Springer Series in Operations Research and Financial Engineering*. Springer International Publishing, Cham.

Powell, W.B., Frantzeskakis, L.F., 1994. Restricted recourse strategies for dynamic networks with random arc capacities. *Transportation Science* 28, 1, 3–23.

Rahimian, H., Bayraksan, G., De-Mello, T.H., 2022. Effective scenarios in multistage distributionally robust optimization with a focus on total variation distance. *SIAM Journal on Optimization* 32, 3, 1698–1727.

Rahimian, H., Bayraksan, G., Homem-de-Mello, T., 2019. Identifying effective scenarios in distributionally robust stochastic programs with total variation distance. *Mathematical Programming* 173, 1-2, 393–430.

Sandıkçı, B., Kong, N., Schaefer, A.J., 2013. A hierarchy of bounds for stochastic mixed-integer programs. *Mathematical Programming* 138, 1-2, 253–272.

Shapiro, A., Dentcheva, D., Ruszczyński, A.P., 2014. Lectures on Stochastic Programming: Modeling and Theory (2nd edn). *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics: Mathematical Optimization Society, Philadelphia, PA.

Shapiro, A., Nemirovski, A., 2005. On complexity of stochastic programming problems. In Jeyakumar, V. and Rubinov, A. (eds), *Continuous Optimization: Current Trends and Modern Applications*. Applied Optimization, Springer US, Boston, MA, pp. 111–146.

Thaler, R.H., 1988. Anomalies: The winner's curse. *Journal of Economic Perspectives* 2, 1, 191–202.

Wallace, S.W., 1987. A piecewise linear upper bound on the network recourse function. *Mathematical Programming* 38, 2, 133–146.

Wallace, S.W., 2010. Stochastic programming and the option of doing it differently. *Annals of Operations Research* 177, 1, 3–8.

Wets, R.J.B., 1984. Modeling and solution strategies for unconstrained stochastic optimization problems. *Annals of Operations Research* 1, 1, 3–22.

Wright, S.E., 1994. Primal-dual aggregation and disaggregation for stochastic linear programs. *Mathematics of Operations Research* 19, 4, 893–908.

Xu, J., Sen, S., 2023. Compromise policy for multi-stage stochastic linear programming: variance and bias reduction. *Computers & Operations Research* 153, 106132.

Zipkin, P.H., 1980a. Bounds for row-aggregation in linear programming. *Operations Research* 28, 4, 903–916.

Zipkin, P.H., 1980b. Bounds on the effect of aggregating variables in linear programs. *Operations Research* 28, 2, 403–418.