CIRRELT

Centre interuniversitaire de recherche
sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre
on Enterprise Networks, Logistics and Transportation

On the Scenario-Tree Optimal-Value Error for Stochastic Programming Problems

Julien Keutchayan
David Munger
Michel Gendreau

December 2017

CIRRELT-2017-73

Bureaux de Montréal :
Université de Montréal
Pavillon André-Aisenstadt
C.P. 6128, succursale Centre-ville
Montréal (Québec)
Canada H3C 3J7
Téléphone : 514 343-7575
Télécopie : 514 343-7121

Bureaux de Québec :
Université Laval
Pavillon Palasis-Prince
2325, de la Terrasse, bureau 2642
Québec (Québec)
Canada G1V 0A6
Téléphone : 418 656-2073
Télécopie : 418 656-2624

www.cirrelt.ca

On the Scenario-Tree Optimal-Value Error for Stochastic Programming Problems[†]

Julien Keutchayan[1,*], David Munger[2,], Michel Gendreau[1]**

[1] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT) and Department of Mathematics and Industrial Engineering, Polytechnique Montréal, P.O. Box 6079, Station Centre-Ville, Montréal, Canada H3C 3A7

[2] Kronos, AD OPT Division, 3535 chemin Queen-Mary, Montréal, Canada, H3V 1H8

Abstract. Stochastic programming problems generally lead to large-scale intractable programs if the number of possible outcomes for the random parameters is large or if the problem has many stages. A way to address these problems at lower computational cost is provided by the scenario-tree generation methods, which construct approximate problems from a finite subset of scenarios. When considering a general problem, the number of scenarios required to keep the optimal-value error within a given range grows exponentially with the number of random parameters and stages, which may lead to approximate problems that are themselves intractable. To overcome this fast growth of complexity, there is a need to look for scenario-tree generation methods tailored to specific classes of problems. In this paper, we provide a theoretical basis to develop such methods by studying the optimal-value error in the context of a general stochastic programming problem. We derive specifically two main results: an error decomposition and an error upper bound, both written as a sum of lower-level errors made at each node of the scenario tree. These two results provide the cornerstone to a new approach that takes into account the structure of the problem to generate suitable scenario trees.

Keywords. Stochastic optimization, multistage stochastic programming, scenario-tree generation, optimal-value error analysis.

Acknowledgements. The authors would like to thank for their support Hydro-Québec and the Natural Sciences and Engineering Research Council of Canada (NSERC) through the NSERC/Hydro-Québec Industrial Research Chair on the Stochastic Optimization of Electricity Generation, Polytechnique Montréal, and the NSERC Discovery Grants Program.

[†] Revised version of CIRRELT-2017-05.

** Part of this research was carried out while David Munger was research professional at CIRRELT and Polytechnique Montréal.

---

* Corresponding author: Julien.Keutchayan@cirrelt.ca

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2017

1 Introduction

Stochastic programming is a particular subarea of mathematical programming used to model optimization problems that involve uncertain parameters. These parameters are modeled by random vectors whose probability distributions are generally inferred from available data. As for the objective function to be maximized, it is often modeled as an expectation of a revenue or utility function; see, e.g., King and Wallace [19]. Stochastic programming problems are found in various fields of optimization, we refer for instance to Yu et al. [41] and Ziemba [42] for applications in finance; Bertocchi et al. [4] and Kovacevic et al. [21] for applications in energy production and trading; Powell and Topaloglu [32] and Yen and Birge [40] for applications in transport and logistic; Beraldi et al. [3] and Colvin and Maravelias [6] for applications in medicine. For a general presentation on the theory and the solution methods in stochastic programming, we refer to Prékopa [33], Ruszczyński and Shapiro [36], and Birge and Louveaux [5].

Stochastic programming problems are generally highly computationally challenging to solve exactly, as shown in Dyer and Stougie [8] and Hanasusanto et al. [10]. A way to address them at smaller computational cost consists in constructing an approximate problem with a finite subset of realizations obtained by discretizing the random parameters. These realizations are called *scenarios* and this solution approach is known as the *scenario-tree generation*. The generation of scenario trees is subject to a trade-off: on the one hand, scenario trees must include a number of scenarios small enough so that the approximate problem can be solved by optimization tools in a reasonable time; on the other hand, this number must be large enough so that the approximate problem estimates accurately the optimal value and solutions of the original problem. This trade-off is fairly easy to satisfy if the problem has a reasonable size; by "size" we mean the number T of stages at which random information is revealed to the decision-maker and the dimension d_t of the random vector at each stage. However, as the size of the problem increases, the trade-off is more and more difficult to manage, and, currently, problems with large T or d_t are typically out-of-scope of scenario-tree generation methods. An important challenge in stochastic programming is therefore the design of efficient scenario trees for large size problems.

Many methods have been proposed to generate scenario trees, we refer in particular to the following works: Shapiro and Homem-de Mello [38] and Shapiro [37] on *Monte Carlo methods*; Pennanen [27], Koivu [20], and Leövey and Römisch [23] on *quasi-Monte Carlo methods* and *numerical integration rules*; Høyland and Wallace [15] and Høyland et al. [16] on *moment matching methods*; Pflug [28] and Pflug and Pichler [31] on *optimal quantization methods*; Dupačová et al. [7], Heitsch and Römisch [11], and Growe-Kuska et al. [9] on *scenario reduction methods*. These methods have their own theoretical or practical justifications. For instance, Monte Carlo and quasi-Monte Carlo methods for two-stage problems are justified by several results about the asymptotic consistency of the methods, which prove that the optimal-value error decreases to zero as the number of scenarios increases to infinity; see, e.g., Shapiro and Homem-de Mello [39], Homem-de Mello [14], Mak et al. [24], and Bastin et al. [1]. The asymptotic consistency of discretization methods in the multistage setting has been studied first by Olsen [25], and more recently by Pennanen [26] who provides conditions under which the optimal value and solutions of the approximate problem converge to those of the original problem (property known as *epi-convergence*). The optimal-value error has also been extensively studied using probability metrics, which measure the distance between the true probability distribution of the random parameters and its scenario-tree approximation sitting on finitely many scenarios; see Pflug and Pichler [29] for a review on probability metrics. Bounds on the optimal-value error by means of probability metrics, also known as *stability results*, have been derived for instance in Heitsch and Römisch [12] and Pflug and Pichler [30]; see also Römisch [35] for a detailed analysis on stability. As for the assessment of scenario-tree generation methods

through the quality of the decisions rather than the optimal-value error, little work has been done in that direction, as it is difficult to quantify the error related to the implementation of the optimal decisions of the approximate and original problems; a general approach to do so is proposed in Keutchayan et al. [17].

As of today, the use of scenario-tree generation methods for problems with many stages or high-dimensional random vectors is limited by the fast growth of the scenario-tree size. We think that this limitation arises because scenario trees are often not suitable for the problem they intend to solve. Indeed, the current methods focus mostly on the discretization quality of the random parameters, with little or no regard to the structure of the problem itself, namely, its objective function and constraints. But it is reasonable to doubt that a particular way to generate scenario trees can suit most problems regardless of their structures. For this reason, we think that it is necessary to identify classes of problems (i.e., problems with similar structure) and generate scenario trees tailored to each class.

The goal of this paper is therefore to provide the ground for a systematic approach to build scenario trees better suited to problems, in order to broaden the class of solvable problems. With that goal in mind, we study the optimal-value error that results from approximately solving a general stochastic programming problem with a scenario tree. We derive specifically two main results: an optimal-value error *decomposition* and an optimal-value error *upper bound*, both written as a sum of lower-level errors made at each node of the scenario tree. The latter errors are called *low-level* because they concern each individual node, as opposed to the optimal-value error which is *high-level* because it concerns the whole scenario tree. Our two main results show how this high-level error emerges as a combination of the low-level errors, and therefore, provide insight into the way the former can be kept small by acting appropriately on the latter. In particular, the upper bound can be viewed as a stability result for multistage problems. However, our result differs from the two main multistage stability results derived in Heitsch et al. [13] and Pflug and Pichler [30] by the fact that we do not mean to bound the optimal-value error by a distance between probability distributions because, in our view, this hides relevant information about the structure of the problem.

The remainder of this paper is organized as follows: Section 2 contains the preliminaries of the two main results; in particular, Section 2.1 introduces the notation for the stochastic programming problem along with five conditions that the problem should satisfy to ensure that the two main results hold; Section 2.2 introduces a more concise notation for the quantities described in Section 2.1, which will simplify the following mathematical developments; and Section 2.3 introduces the notation for the scenario tree and the approximate problem. Sections 3 and 4 contain the error decomposition and the error bound, respectively, and Section 5 shows how these results can be used to improve scenario-tree generation. Finally, Section 6 concludes the paper.

2 Preliminaries

We consider a *stochastic programming problem* where decisions are made at discrete time stages $t = 0, 1, \ldots, T \in \mathbb{N}_+$, where $\mathbb{N}_+$ stands for the positive integers. *Multistage* problems correspond to the case $T \geq 2$, while *two-stage* problems correspond to $T = 1$. For the sake of conciseness, all results in this paper are formulated for $T \geq 2$, but the reader can easily deduce the corresponding results for two-stage problems.

2.1 Stochastic programming problem formulation

Stochastic programming problems deal with random parameters that are represented by a discrete-time stochastic process of the form $(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_T)$, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Each

random vector $\boldsymbol{\xi}_t$ contains the random parameters revealed during the period $(t-1, t)$, its probability distribution has a support $\Xi_t \subseteq \mathbb{R}^{d_t}$ with $d_t \in \mathbb{N}_+$. Throughout this paper, random vectors are distinguished from their realizations by writing the former in bold font. We denote by $\Xi_{..t}$ the support of $\boldsymbol{\xi}_{..t} := (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_t)$, and by $\Xi_t(\xi_{..t-1})$ the conditional support of $\boldsymbol{\xi}_t$ given the event $\{\omega \in \Omega : \boldsymbol{\xi}_{..t-1}(\omega) = \xi_{..t-1}\} \in \mathcal{A}$. We emphasize that no specific form of probability distribution is assumed for the stochastic process, hence the latter can be discrete or continuous (or a combination of both).

The decision vector y_t at stage $t \in \{0, \ldots, T\}$ satisfies the constraints of the problem if it belongs to the *feasible set* $Y_t \subseteq \mathbb{R}^{s_t}$ with $s_t \in \mathbb{N}_+$. For the sake of clarity, and without loss of generality, we assume throughout this paper that $s_t = s$ and $d_t = d$ for all t. When $t \geq 1$, we also denote the feasible set by $Y_t(y_{..t-1}; \xi_{..t})$ to emphasize that it possibly depends on the decisions $y_{..t-1} := (y_0, \ldots, y_{t-1}) \in \mathbb{R}^{st}$ and on the realization $\xi_{..t} \in \Xi_{..t}$ up to stage t. We consider specifically feasible sets represented as the solutions of finitely or countably many equality and inequality constraints:

Condition 1. (i) The decision vector y_0 belongs to Y_0 if and only if y_0 satisfies $g_{0,i}(y_0) = 0$ for $i \in I_0$ and $g_{0,i}(y_0) \geq 0$ for $i \in J_0$, where $g_{0,i} : \mathbb{R}^s \to \mathbb{R}$ is continuous for $i \in I_0$ and upper semi-continuous for $i \in J_0$ and I_0, J_0 are some finite or countable index sets. (ii) For each $t \in \{1, \ldots, T\}$ and $\xi_{..t} \in \Xi_{..t}$, define $Z_t(\xi_{..t}) := \{y_{..t} \in \mathbb{R}^{s(t+1)} : y_{..t-1} \in Z_{t-1}(\xi_{..t-1}),\ y_t \in Y_t(y_{..t-1}; \xi_{..t})\}$, where $Z_0 := Y_0$. The decision vector $y_{..t}$ belongs to $Z_t(\xi_{..t})$ if and only if $y_{..t}$ satisfies $g_{t,i}(y_{..t}; \xi_{..t}) = 0$ for $i \in I_t$ and $g_{t,i}(y_{..t}; \xi_{..t}) \geq 0$ for $i \in J_t$, where $g_{t,i} : \mathbb{R}^{s(t+1)} \times \Xi_{..t} \to \mathbb{R}$ is a Carathéodory integrand for $i \in I_t$ and a normal integrand for $i \in J_t$ and I_t, J_t are some finite or countable index sets.

The above functions $g_{t,i}$, for $t \in \{0, \ldots, T\}$ and $i \in I_t \cup J_t$, model the constraints of the stochastic programming problem. For the definitions of *normal* and *Carathéodory* integrands, we refer to Rockafellar and Wets [34], Definition 14.27 and Example 14.29, respectively; note that, since in this paper we consider a maximization problem, normal integrands are *upper* semi-continuous by convention. The motivation for Condition 1 is the fact that it implies that the set Z_0 is closed and that the set-valued mapping $\xi_{..t} \rightrightarrows Z_t(\xi_{..t})$ is closed-valued and measurable; see [34, Theorem 14.36].

We also require a boundedness condition on Z_0 and $Z_t(\xi_{..t})$, which, together with Condition 1, imply that Z_0 and $Z_t(\xi_{..t})$ are compact sets for every $\xi_{..t}$.

Condition 2. The set Z_0 is bounded in $\mathbb{R}^s$ and so is $Z_t(\xi_{..t})$ in $\mathbb{R}^{s(t+1)}$ for every $t \in \{1, \ldots, T\}$ and $\xi_{..t} \in \Xi_{..t}$.

We restrict our attention to stochastic programming problems that have a non-empty feasible set and a relative complete recourse at every stage, as expressed in Condition 3:

Condition 3. The set Y_0 is non-empty and so is $Y_t(y_{..t-1}; \xi_{..t})$ for every $t \in \{1, \ldots, T\}$, $\xi_{..t} \in \Xi_{..t}$, and $y_{..t-1} \in Z_{t-1}(\xi_{..t-1})$.

We introduce a *revenue function* $q : \mathbb{R}^{s(T+1)} \times \Xi_{..T} \to \mathbb{R}$. The value $q(y_{..T}; \xi_{..T})$ represents all the revenues obtained from stage 0 to T with the decisions $y_{..T} = (y_0, \ldots, y_T)$ in the realization $\xi_{..T}$. Given the stochastic process, the feasible sets, and the revenue function, we can now define the stage-t *expected recourse function* $\widetilde{Q}_t$ and *optimal recourse function* $\widetilde{Q}_t^*$, which characterize the stochastic programming problem. These functions are defined recursively by the following stochastic dynamic programming equations:

$$\widetilde{Q}_t^*(y_{..t-1}; \xi_{..t}) := \sup_{y_t \in \mathbb{R}^s :\, (y_{..t-1}, y_t) \in Z_t(\xi_{..t})} \widetilde{Q}_t(y_{..t}; \xi_{..t}), \quad \forall t \in \{0, \ldots, T\}, \tag{1}$$

$$\widetilde{Q}_t(y_{..t}; \xi_{..t}) := \mathbb{E}[\widetilde{Q}_{t+1}^*(y_{..t}; \xi_{..t}, \boldsymbol{\xi}_{t+1}) \,|\, \boldsymbol{\xi}_{..t} = \xi_{..t}], \quad \forall t \in \{0, \ldots, T-1\}, \tag{2}$$

where at $t = T$ the equation (1) is initialized by setting

$$\widetilde{Q}_T(y_{..T}; \xi_{..T}) := q(y_{..T}; \xi_{..T}), \quad \forall \xi_{..T} \in \Xi_{..T}, \forall y_{..T} \in \mathbb{R}^{s(T+1)}, \tag{3}$$

and at $t = 0$ the arguments $y_{..t-1}$ and $\xi_{..t}$ are removed, i.e.,

$$\widetilde{Q}_0(y_{..0}; \xi_{..0}) = \widetilde{Q}_0(y_0) \quad \text{and} \quad \widetilde{Q}_0^*(y_{..0-1}; \xi_{..0}) = \widetilde{Q}_0^*. \tag{4}$$

To ensure that the (expected and optimal) recourse functions are well-defined, we add the following two conditions on the revenue function.

Condition 4. The function $q(\cdot; \xi_{..T})$ is upper semi-continuous for every $\xi_{..T} \in \Xi_{..T}$ and $q(\cdot; \cdot)$ is $\mathcal{B}(\mathbb{R}^{s(T+1)}) \otimes \Sigma$-measurable, where $\mathcal{B}(\mathbb{R}^{s(T+1)})$ is the Borel σ-algebra of $\mathbb{R}^{s(T+1)}$ and $(\Xi_{..T}, \Sigma)$ is a complete measurable space with respect to the distribution of $\boldsymbol{\xi}_{..T}$.

Condition 4 directly implies that the revenue function is a normal integrand by [34, Corollary 14.34].

Condition 5. There exists a measurable function $h : \Xi_{..T} \to \mathbb{R}$ such that $|q(y_{..T}; \xi_{..T})| \leq h(\xi_{..T})$ for all $\xi_{..T} \in \Xi_{..T}$ and $y_{..T} \in Z_T(\xi_{..T})$, where h satisfies $\mathbb{E}[|h(\boldsymbol{\xi}_{..T})|] < \infty$ and $\mathbb{E}[|h(\boldsymbol{\xi}_{..T})| \,|\, \boldsymbol{\xi}_{..t} = \xi_{..t}] < \infty$ for all $t \in \{1, \ldots, T-1\}$ and $\xi_{..t} \in \Xi_{..t}$.

Note that Condition 5 requires that the conditional integrability of $h(\boldsymbol{\xi}_{..T})$ given $\boldsymbol{\xi}_{..t} = \xi_{..t}$ holds for any $\xi_{..t} \in \Xi_{..t}$, and not merely for almost every $\xi_{..t} \in \Xi_{..t}$. The reason is that we want the stage-t (expected and optimal) recourse functions to be defined everywhere on $\Xi_{..t}$, which will guarantee that the *node errors* and the *subtree errors*, introduced in Section 3, are well-defined even if the scenarios are chosen in a non-random fashion.

We shall show now that the five conditions above guarantee the existence of optimal decision vectors at every stage and the finiteness of the (expected and optimal) recourse functions. We do so by proving recursively, from stage T to 0, the existence of optimal solutions for the optimization problem at the right-hand side of (1). In the following, we denote by $\delta_C(\cdot)$ the function defined as $\delta_C(x) = 0$ if $x \in C$ and $\delta_C(x) = -\infty$ otherwise. Through this notation, we can express the fact that the supremum of a real-valued function f over a set $C \subseteq \mathbb{R}^s$ is written equivalently as the supremum of the *extended*-real-valued function $f + \delta_C$ over $\mathbb{R}^s$; see [34, Chapter 1] for detailed developments on extended real analysis.

Take an arbitrary $t \in \{1, \ldots, T\}$ and suppose that the stage-t expected recourse function $(y_{..t}; \xi_{..t}) \mapsto \widetilde{Q}_t(y_{..t}; \xi_{..t})$ is a normal integrand and that its effective domain

$$\operatorname{dom} \widetilde{Q}_t(\cdot\,; \xi_{..t}) := \{y_{..t} \in \mathbb{R}^{s(t+1)} : \widetilde{Q}_t(y_{..t}; \xi_{..t}) > -\infty\}, \tag{5}$$

includes $Z_t(\xi_{..t})$ for every $\xi_{..t} \in \Xi_{..t}$.

It follows from the above induction hypothesis and the properties of measurability and compactness of $Z_t(\xi_{..t})$ that the mapping

$$(y_{..t}; \xi_{..t}) \mapsto \widetilde{Q}_t(y_{..t}; \xi_{..t}) + \delta_{Z_t(\xi_{..t})}(y_{..t}), \tag{6}$$

is a normal integrand too and, moreover, is level-bounded in y_t locally uniformly in $y_{..t-1}$ for each fixed $\xi_{..t} \in \Xi_{..t}$; see [34, Example 14.32, Definition 1.16]. Thus, the stage-t optimal recourse function, defined as

$$\widetilde{Q}_t^*(y_{..t-1}; \xi_{..t}) = \sup_{y_t \in \mathbb{R}^s} \{\widetilde{Q}_t(y_{..t-1}, y_t; \xi_{..t}) + \delta_{Z_t(\xi_{..t})}(y_{..t-1}, y_t)\}, \tag{7}$$

is a normal integrand by [34, Proposition 14.47]. Moreover, take an arbitrary $\xi_{..t} \in \Xi_{..t}$ and consider the following two cases: (i) if $y_{..t-1} \in Z_{t-1}(\xi_{..t-1})$, then $Z_t(\xi_{..t}) \neq \emptyset$ by Condition 3, and hence the supremum in (7) is attained, $\widetilde{Q}_t^*(y_{..t-1}; \xi_{..t})$ is finite, and an optimal solution $y_t^* =: x_t^*(y_{..t-1}; \xi_{..t})$ exists, where we introduce the notation $x_t^*(y_{..t-1}; \xi_{..t})$ to emphasize that this optimal solution depends on $y_{..t-1}$ and $\xi_{..t}$; (ii) if $y_{..t-1} \notin Z_{t-1}(\xi_{..t-1})$, then the supremum in (7) equals $-\infty$, and this value is consistent with the fact that such $y_{..t-1}$ is not a vector of feasible decisions. Therefore, for every $\xi_{..t} \in \Xi_{..t}$ we have that

$$\widetilde{Q}_t^*(y_{..t-1}; \xi_{..t}) \begin{cases} \in \mathbb{R} & \text{if } y_{..t-1} \in Z_{t-1}(\xi_{..t-1}); \\ = -\infty & \text{otherwise.} \end{cases} \tag{8}$$

This concludes the analysis of the equation (1).

As for the equation (2), we shall prove that the stage-$(t-1)$ expected recourse function

$$\widetilde{Q}_{t-1}(y_{..t-1}; \xi_{..t-1}) = \mathbb{E}[\widetilde{Q}_t^*(y_{..t-1}; \xi_{..t-1}, \boldsymbol{\xi}_t) \,|\, \boldsymbol{\xi}_{..t-1} = \xi_{..t-1}], \tag{9}$$

is a normal integrand and that $\operatorname{dom} \widetilde{Q}_{t-1}(\cdot\,; \xi_{..t-1})$ includes $Z_{t-1}(\xi_{..t-1})$ for every $\xi_{..t-1} \in \Xi_{..t-1}$. This will allow the above arguments to be repeated at stage $t-1$, and hence will complete the proof since the initial stage-T expected recourse function is a finite-valued normal integrand by (3) and Condition 4. Take an arbitrary $\xi_{..t-1} \in \Xi_{..t-1}$ and consider the following two cases: (i) if $y_{..t-1} \in Z_{t-1}(\xi_{..t-1})$, then it follows from Condition 5 and an application of Lebesgue's dominated convergence theorem that $\widetilde{Q}_{t-1}(\,\cdot\,; \xi_{..t-1})$ is finite-valued and upper semi-continuous at $y_{..t-1}$; (ii) if $y_{..t-1} \notin Z_{t-1}(\xi_{..t-1})$, then we have by (8) that $\widetilde{Q}_t^*(y_{..t-1}; \xi_{..t-1}, \xi_t) = -\infty$ for all $\xi_t \in \Xi_t(\xi_{..t-1})$, and hence $\widetilde{Q}_{t-1}(y_{..t-1}; \xi_{..t-1}) = -\infty$. Since $\xi_{..t-1} \rightrightarrows Z_{t-1}(\xi_{..t-1})$ is closed-valued and $\widetilde{Q}_{t-1}$ remains measurable, we deduce from (i)-(ii) that $\widetilde{Q}_{t-1}$ is a normal integrand and that its effective domain satisfies $\operatorname{dom} \widetilde{Q}_{t-1}(\cdot\,; \xi_{..t-1}) = Z_{t-1}(\xi_{..t-1})$ for every $\xi_{..t-1} \in \Xi_{..t-1}$; see [34, Corollary 14.34].

To summarize, the above arguments prove that for every $t \in \{0, \ldots, T\}$, $\xi_{..t} \in \Xi_{..t}$, and $y_{..t} \in Z_t(\xi_{..t})$, an optimal decision $x_t^*(y_{..t-1}; \xi_{..t})$ exists and the values $\widetilde{Q}_t^*(y_{..t-1}; \xi_{..t})$ and $\widetilde{Q}_t(y_{..t}; \xi_{..t})$ are finite (with the appropriate change of arguments for $t = 0$).

2.2 Decision policy formulation

It is more convenient for the future developments in this paper to introduce a single notation for the stage-t expected and optimal recourse functions defined in (1)-(2). In Section 2.1, we have introduced $x_t^*(y_{..t-1}; \xi_{..t})$ to denote the optimal decision vector at stage t given as a function of the decisions $y_{..t-1}$ and the realization $\xi_{..t}$ up to t. We can generalize this notation to represent *any* feasible decision at stage t as a function of $(y_{..t-1}; \xi_{..t})$. The development below formalizes this approach and shows the link with the previous formulation.

We model a decision policy $x_{..T} := (x_0, \ldots, x_T)$ as a collection of a decision vector $x_0 \in \mathbb{R}^s$ and several decision functions x_1, ..., x_T such that the value $x_t(y_{..t-1}; \xi_{..t})$ specifies the decision at stage t as a function of $y_{..t-1}$ and $\xi_{..t}$. The fact that x_t does not depend on the realizations after stage t ensures that the policy is non-anticipative.

The feasibility constraints are now modeled using subsets of function spaces. Specifically, we define for each $t \in \{1, \ldots, T\}$ the set $\mathcal{X}_t$ of all stage-t feasible decision functions for the stochastic programming problem as

$$\mathcal{X}_1 = \{x_1 : Z_0 \times \Xi_1 \to \mathbb{R}^s : \forall \xi_1 \in \Xi_1,\, \forall y_0 \in Z_0,\, x_1(y_0; \xi_1) \in Y_1(y_0; \xi_1)\}, \tag{10}$$

and for each $t \in \{2, \ldots, T\}$,

$$\mathcal{X}_t = \big\{x_t : \bigcup_{\xi_{..t} \in \Xi_{..t}} (Z_{t-1}(\xi_{..t-1}) \times \{\xi_{..t}\}) \to \mathbb{R}^s : \\ \forall \xi_{..t} \in \Xi_{..t},\, \forall y_{..t-1} \in Z_{t-1}(\xi_{..t-1}),\, x_t(y_{..t-1}; \xi_{..t}) \in Y_t(y_{..t-1}; \xi_{..t})\big\}. \tag{11}$$

At stage 0, we define $\mathcal{X}_0 = Z_0$, as x_0 is a vector and not a function. The set $\mathcal{X}_{..t}$ of all feasible decision policies $x_{..t} = (x_0, \ldots, x_t)$ for the stochastic programming problem up to stage t is given by $\mathcal{X}_{..t} := \Pi_{i=0}^{t} \mathcal{X}_i$, where $\Pi_{i=0}^{t}$ denotes the $(t+1)$-fold Cartesian product.

The connection between decision *policies* and decision *vectors* is as follows: for a policy $x_{..T} \in \mathcal{X}_{..T}$ and a realization $\xi_{..T} \in \Xi_{..T}$, the associated vector $y_{..T} = (y_0, \ldots, y_T)$ is given by

$$y_t = \begin{cases} x_0 & \text{if } t = 0; \\ x_t(y_{..t-1}; \xi_{..t}) & \text{if } t \in \{1, \ldots, T\}. \end{cases} \tag{12}$$

We introduce the stage-t *(generalized) recourse function* $Q_t : \mathcal{X}_{..T} \times \Xi_{..t} \to \mathbb{R}$ for every $t \in \{1, \ldots, T\}$ and $Q_0 : \mathcal{X}_{..T} \to \mathbb{R}$ at stage 0. The value $Q_t(x_{..T}; \xi_{..t})$ represents the conditional expectation of revenues obtained by implementing the policy $x_{..T} \in \mathcal{X}_{..T}$ given the realization $\xi_{..t} \in \Xi_{..t}$ up to stage t. The recourse functions are computed by backward recursion from stage T to 0 by the relation

$$Q_t(x_{..T}; \xi_{..t}) := \mathbb{E}\big[Q_{t+1}(x_{..T}; \boldsymbol{\xi}_{..t+1}) \,\big|\, \boldsymbol{\xi}_{..t} = \xi_{..t}\big], \quad \forall x_{..T} \in \mathcal{X}_{..T},\, \forall \xi_{..t} \in \Xi_{..t}, \tag{13}$$

where at stage T the relation is initialized by setting

$$Q_T(x_{..T}; \xi_{..T}) := q(y_{..T}; \xi_{..T}), \quad \text{with } y_{..T} \text{ given by (12)}, \tag{14}$$

and at $t = 0$ the argument $\xi_{..t}$ is removed, i.e., $Q_0(x_{..T}) = \mathbb{E}\big[Q_1(x_{..T}; \boldsymbol{\xi}_1)\big]$.

In this setting, an *optimal decision policy* for the stochastic programming problem is a policy $x^*_{..T} = (x^*_0, \ldots, x^*_T) \in \mathcal{X}_{..T}$ such that the following holds:

- at stage T:

$$Q_T(x_{..T-1}, x^*_T; \xi_{..T}) \geq Q_T(x_{..T-1}, x_T; \xi_{..T}), \quad \forall \xi_{..T} \in \Xi_{..T},\, \forall x_{..T} \in \mathcal{X}_{..T}; \tag{15}$$

- at stage $t \in \{1, \ldots, T-1\}$:

$$Q_t(x_{..t-1}, x^*_t, x^*_{t+1..}; \xi_{..t}) \geq Q_t(x_{..t-1}, x_t, x^*_{t+1..}; \xi_{..t}), \quad \forall \xi_{..t} \in \Xi_{..t},\, \forall x_{..t} \in \mathcal{X}_{..t}; \tag{16}$$

- at stage 0:

$$Q_0(x^*_0, x^*_{1..}) \geq Q_0(x_0, x^*_{1..}), \quad \forall x_0 \in \mathcal{X}_0, \tag{17}$$

where we use the shorthand $x^*_{t..} := (x^*_t, \ldots, x^*_T)$. The quantity $Q_0(x^*_{..T})$ is the *optimal value* of the stochastic programming problem.

The inequalities (15)-(17) mean intuitively the following: when one uses $x^*_{t+1..}$ to make decisions from stage $t+1$ to T, the stage-t decision function x^*_t is optimal for the recourse function $x_t \mapsto Q_t(x_{..t-1}, x_t, x^*_{t+1..}; \xi_{..t})$ regardless of the policy $x_{..t-1} \in \mathcal{X}_{..t-1}$ used to make decisions from stage 0 to $t-1$. This is Bellman's principle of optimality; see Bellman [2].

It follows from the five conditions introduced in Section 2.1 that both sides of the inequalities (15)-(17) are well-defined and finite-valued for any feasible policy and random realization.

2.3 Scenario-tree and approximate problem formulations

The optimal value $Q_0(x^*)$ and the optimal policy x^* of the stochastic programming problem are not readily available in general. The scenario-tree approach to estimate Q_t consists in approximating the right-hand side of (13) as a weighted average of the values of Q_{t+1} for a selection of realizations of $\boldsymbol{\xi}_{t+1}$. In turn, Q_{t+1} is approximated in terms of Q_{t+2}, and this recursive discretization scheme is carried out to stage T, where the values of Q_T are computed directly from the revenue function q. A tree structure naturally arises from this scheme, in which sibling nodes at stage $t+1$ represent the discrete values of Q_{t+1} whose weighted average approximates the value of Q_t, represented by their common parent node at stage t. The remainder of this section formalizes the scenario tree and the approximate problem.

A scenario tree is a *rooted tree structure* $\mathcal{T} = (\mathcal{N}, \mathcal{E}, n_0)$, with $\mathcal{N}$ the (finite) node set, $\mathcal{E}$ the edge set and n_0 the root node. This structure is such that T edges separate the root from any of the tree leaves, where a *leaf* is any node $n \in \mathcal{N}^* := \mathcal{N} \setminus \{n_0\}$ incident to only one edge. We introduce the notation $C(n)$, $p(n)$, and $t(n)$ to denote, respectively, the set of *children nodes* of n (i.e., the nodes linked to n at the next stage), the *parent node* of n (i.e., the node linked to n at the previous stage), and the *stage* (or *depth*) of n (i.e., the number of edges between n_0 and n). We denote by $[n_0, n]$ the unique sequence of nodes from n_0 to n (we write $(n_0, m]$ when n_0 is excluded from that sequence) and $\mathcal{N}_t := \{n \in \mathcal{N} : t(n) = t\}$.

Every node $n \in \mathcal{N}^*$ carries a *discretization weight* w^n and a *discretization point* ζ^n of $\boldsymbol{\xi}_{t(n)}$, which satisfy

$$w^n > 0 \quad \text{and} \quad \zeta^n \in \begin{cases} \Xi_1 & \text{if } n \in \mathcal{N}_1; \\ \Xi_{t(n)}(\zeta^{\cdot\cdot p(n)}) & \text{if } n \in \mathcal{N}^* \setminus \mathcal{N}_1, \end{cases} \tag{18}$$

where the notation $\zeta^{\cdot\cdot n}$, defined for each $n \in \mathcal{N}^*$, refers to the sequence of discretization points from n_0 to n, i.e., $\zeta^{\cdot\cdot n} := (\zeta^m)_{m \in (n_0, n]}$. The value w^n represents the weight of node n with respect to its sibling nodes, i.e., $C(p(n))$. The weight of n with respect to all the stage-$t(n)$ nodes $\mathcal{N}_{t(n)}$ is given by the *product weight* W^n defined as

$$W^n = \begin{cases} 1 & \text{if } n = n_0; \\ \prod_{m \in (n_0, n]} w^n & \text{if } n \in \mathcal{N}^*. \end{cases} \tag{19}$$

Now that we have formalized the scenario tree, we can define the approximate problem that corresponds to it. To this end, we define for each stage $t \in \{1, \dots, T\}$ the set $\widehat{\mathcal{X}}_t$ of all stage-t feasible decision functions for the approximate problem by

$$\widehat{\mathcal{X}}_1 = \big\{x_1 : Z_0 \times \{\zeta^n : n \in \mathcal{N}_1\} \to \mathbb{R}^s : \forall n \in \mathcal{N}_1,\, \forall y_0 \in Z_0,\, x_1(y_0; \zeta^n) \in Y_1(y_0; \zeta^n)\big\}, \tag{20}$$

and for each $t \in \{2, \dots, T\}$,

$$\begin{aligned} \widehat{\mathcal{X}}_t = \big\{x_t : \bigcup_{n \in \mathcal{N}_t} (Z_{t-1}(\zeta^{\cdot\cdot p(n)}) \times \{\zeta^{\cdot\cdot n}\}) \to \mathbb{R}^s : \\ \forall n \in \mathcal{N}_t,\, \forall y_{\cdot\cdot t-1} \in Z_{t-1}(\zeta^{\cdot\cdot p(n)}),\, x_t(y_{\cdot\cdot t-1}; \zeta^{\cdot\cdot n}) \in Y_t(y_{\cdot\cdot t-1}; \zeta^{\cdot\cdot n})\big\}. \end{aligned} \tag{21}$$

At stage 0, we define $\widehat{\mathcal{X}}_0 = Z_0$ (hence $\widehat{\mathcal{X}}_0 = \mathcal{X}_0$). The set $\widehat{\mathcal{X}}_{\cdot\cdot t}$ of all feasible decision policies $\widehat{x}_{\cdot\cdot t} = (\widehat{x}_0, \dots, \widehat{x}_t)$ for the approximate problem up to stage t is $\widehat{\mathcal{X}}_{\cdot\cdot t} := \Pi_{i=0}^t \widehat{\mathcal{X}}_i$.

It should be noted that in a general setting there is no inclusion relation between $\mathcal{X}_t$ and $\widehat{\mathcal{X}}_t$, because $\mathcal{X}_t$ contains functions defined on $\Xi_{\cdot\cdot t}$, whereas $\widehat{\mathcal{X}}_t$ contains functions defined on $\{\zeta^{\cdot\cdot n} : n \in \mathcal{N}_t\}$, and the latter is a strict subset of the former whenever the scenario tree does not include

all realizations of the stochastic process up to stage t. It is also important to note that a policy $x_{..T} \in \mathcal{X}_{..T}$ of the original problem carries more information than a policy $\widehat{x}_{..T} \in \widehat{\mathcal{X}}_{..T}$ of the approximate problem. Indeed, one can use $x_{..T}$ to make decisions in the approximate problem, but one cannot use $\widehat{x}_{..T}$ to make decisions in the original problem. This is true in general, however, a subtlety arises when the stage-t realization $\xi_{..t}$ coincides with a discretization sequence $\zeta^{..n}$ for some $n \in \mathcal{N}_t$. In this case, any policy $\widehat{x}_{..t} \in \widehat{\mathcal{X}}_{..t}$ can be used to make decisions from stage 0 to t in the approximate and original problems. For this reason, we extend the domain of definition of the stage-t recourse function $Q_t(\,\cdot\,;\xi_{..t})$ to include this particular case. The new definition is

$$Q_t(\,\cdot\,;\xi_{..t}) : \begin{cases} \Pi_{i=0}^{t}(\mathcal{X}_i \cup \widehat{\mathcal{X}}_i) \times \Pi_{i=t+1}^{T}\mathcal{X}_i \to \mathbb{R} & \text{if } \xi_{..t} = \zeta^{..n} \text{ for some } n \in \mathcal{N}_t; \\ \mathcal{X}_{..T} \to \mathbb{R} & \text{otherwise.} \end{cases} \tag{22}$$

At stage 0, we still have $Q_0 : \mathcal{X}_{..T} \to \mathbb{R}$.

The scenario tree provides at each node an estimator for the recourse function (13). The node-n estimator of the stage-$t(n)$ recourse function $Q_{t(n)}(x_{..T};\zeta^{..n})$ is denoted by $\widehat{Q}^n(x_{..T})$ and is computed recursively from the estimators at node $m \in C(n)$ by

$$\widehat{Q}^n(x_{..T}) = \sum_{m \in C(n)} w^m\, \widehat{Q}^m(x_{..T}), \quad \forall n \in \mathcal{N} \setminus \mathcal{N}_T,\, \forall x_{..T} \in \Pi_{t=0}^{T}(\mathcal{X}_t \cup \widehat{\mathcal{X}}_t), \tag{23}$$

where at node $n \in \mathcal{N}_T$ the relation is initialized by setting

$$\widehat{Q}^n(x_{..T}) = q(y_{..T};\zeta^{..n}), \quad \text{with } y_{..T} \text{ given by (12) in the scenario } \zeta^{..n}. \tag{24}$$

We emphasize that our formulation of the scenario tree estimators is general, since we do not assume a specific form for the tree structure and for the discretization points and weights. This generality allows us to consider essentially all ways to generate scenario trees. For instance, the fact that the weights w^m in (23) need not sum to one can account for the use of *importance sampling* or *integration rules* methods; see, e.g., Shapiro [37] and Pennanen [27], respectively. A well-known particular case of the scenario-tree estimator is the so-called *sample average approximation*, which is obtained from (23) by setting $w^m = |C(n)|^{-1}$ and by getting ζ^m through Monte Carlo methods.

The optimal decision policy of the approximate problem, denoted by $\widehat{x}^*_{..T} = (\widehat{x}^*_0, \ldots, \widehat{x}^*_T) \in \widehat{\mathcal{X}}_{..T}$, satisfies the discrete counterpart of (15)-(17):

- at stage T:

$$\widehat{Q}^n(x_{..T-1}, \widehat{x}^*_T) \geq \widehat{Q}^n(x_{..T-1}, x_T), \quad \forall n \in \mathcal{N}_T,\, \forall x_{..T} \in \Pi_{t=0}^{T}(\mathcal{X}_t \cup \widehat{\mathcal{X}}_t); \tag{25}$$

- at stage $t \in \{1, \ldots, T-1\}$:

$$\widehat{Q}^n(x_{..t-1}, \widehat{x}^*_t, \widehat{x}^*_{t+1..}) \geq \widehat{Q}^n(x_{..t-1}, x_t, \widehat{x}^*_{t+1..}), \quad \forall n \in \mathcal{N}_t,\, \forall x_{..t} \in \Pi_{i=0}^{t}(\mathcal{X}_i \cup \widehat{\mathcal{X}}_i); \tag{26}$$

- at stage 0:

$$\widehat{Q}^{n_0}(\widehat{x}^*_0, \widehat{x}^*_{1..}) \geq \widehat{Q}^{n_0}(x_0, \widehat{x}^*_{1..}), \quad \forall x_0 \in \widehat{\mathcal{X}}_0. \tag{27}$$

The quantity $\widehat{Q}^{n_0}(\widehat{x}^*_{..T})$ is the optimal value of the approximate problem; it is the estimator of $Q_0(x^*_{..T})$ and the quantity $Q_0(x^*_{..T}) - \widehat{Q}^{n_0}(\widehat{x}^*_{..T})$ is what we refer to as the optimal-value error. It follows from the five conditions in Section 2.1 that an optimal decision policy exists and that both sides of (25)-(27) are well-defined and finite-valued for any feasible decision policy and any node in the scenario tree.

We end this section by a remark on two cases of equality between the recourse functions and their estimators.

Remark 2.1. Since the stage-T recourse function Q_T and its estimator $\widehat{Q}^n$ at any node $n \in \mathcal{N}_T$ are both computed directly from the revenue function q, we have that

$$\widehat{Q}^n(x_{..T}) = Q_T(x_{..T}; \zeta^{..n}), \quad \forall n \in \mathcal{N}_T, \forall x_{..T} \in \Pi_{t=0}^{T}(\mathcal{X}_t \cup \widehat{\mathcal{X}}_t). \tag{28}$$

Another case of equality is obtained by noticing that the inequality (15) applied at $\xi = \zeta^{..n}$, for any $n \in \mathcal{N}_T$, provides the same optimality condition than (25). Consequently, the decision functions $\widehat{x}_T^*(\cdot; \zeta^{..n})$ and $x_T^*(\cdot; \zeta^{..n})$ can be chosen such that they coincide, and hence

$$\widehat{Q}^n(x_{..T-1}, \widehat{x}_T^*) = Q_T(x_{..T-1}, x_T^*; \zeta^{..n}), \quad \forall n \in \mathcal{N}_T, \forall x_{..T-1} \in \Pi_{t=0}^{T-1}(\mathcal{X}_t \cup \widehat{\mathcal{X}}_t). \tag{29}$$

3 Node-by-node decomposition of the optimal-value error

The main result of this section is Theorem 3.5, which provides a node-by-node decomposition of the optimal-value error $Q_0(x_{..T}^*) - \widehat{Q}^{n_0}(\widehat{x}_{..T}^*)$.

We start by introducing the concepts of (low-level) *node errors* and (high-level) *subtree errors*. It appears from the stochastic dynamic programming equations (1)-(2) that the optimal-value error results from successive errors made by approximating alternatively the right-hand side of (1) and (2). We call specifically *node optimization error* the error made by approximating (1) and *node discretization error* the error made by approximating (2), at a particular node in the scenario-tree. Their explicit definitions below are expressed by means of the decision policy formulation of Section 2.2:

Definition 3.1 (Node optimization error)**.** *For each stage $t \in \{1, \ldots, T-1\}$, we define the* optimization error $\mathrm{E}_{\mathrm{opt}}^n(x_{..t-1})$ *at node $n \in \mathcal{N}_t$ and for a decision policy $x_{..t-1} \in \Pi_{i=0}^{t-1}(\mathcal{X}_i \cup \widehat{\mathcal{X}}_i)$ as*

$$\mathrm{E}_{\mathrm{opt}}^n(x_{..t-1}) = Q_t(x_{..t-1}, x_t^*, x_{t+1..}^*; \zeta^{..n}) - Q_t(x_{..t-1}, \widehat{x}_t^*, x_{t+1..}^*; \zeta^{..n}). \tag{30}$$

At the root node, the optimization error *is*

$$\mathrm{E}_{\mathrm{opt}}^{n_0} = Q_0(x_0^*, x_{1..}^*) - Q_0(\widehat{x}_0^*, x_{1..}^*). \tag{31}$$

It follows from the optimality conditions (15)-(17) that the optimization errors are always non-negative. The node-n optimization error measures the error at stage $t(n)$ made by using the optimal decision function $\widehat{x}_{t(n)}^*$ of the approximate problem instead of the optimal decision function $x_{t(n)}^*$ of the original problem. The optimization error is not defined at stage T because no optimization error is made at $n \in \mathcal{N}_T$ (cf. Remark 2.1).

Definition 3.2 (Node discretization error)**.** *For each stage $t \in \{0, \ldots, T-1\}$, we define the* discretization error $\mathrm{E}_{\mathrm{disc}}^n(x_{..t})$ *at node $n \in \mathcal{N}_t$ and for a decision policy $x_{..t} \in \Pi_{i=0}^{t}(\mathcal{X}_i \cup \widehat{\mathcal{X}}_i)$ as*

$$\mathrm{E}_{\mathrm{disc}}^n(x_{..t}) = Q_t(x_{..t}, x_{t+1..}^*; \zeta^{..n}) - \sum_{m \in C(n)} w^m \, Q_{t+1}(x_{..t}, x_{t+1..}^*; \zeta^{..m}), \tag{32}$$

if $t \in \{1, \ldots, T-1\}$, and

$$\mathrm{E}_{\mathrm{disc}}^{n_0}(x_0) = Q_0(x_0, x_{1..}^*) - \sum_{m \in C(n_0)} w^m \, Q_1(x_0, x_{1..}^*; \zeta^m), \tag{33}$$

at the root node.

The node-n discretization error measures the error made by substituting the conditional expectation by a finite sum over the children nodes of n.

We define now the concept of *subtree errors*. We call *subtree rooted at node* $n \in \mathcal{N}$ the scenario tree $(\mathcal{N}(n), \mathcal{E}(n), n)$ obtained by setting n as the root node and by considering only the nodes that are the descendants of n and the edges connecting them. The subtree rooted at n_0 is the whole scenario tree $(\mathcal{N}, \mathcal{E}, n_0)$. We distinguish between two subtree errors: the *optimal subtree error* measures the error between the optimal recourse function (1) and its scenario-tree estimator, whereas the *suboptimal subtree error* measures the error between the expected recourse function (2) and its scenario-tree estimator. Their explicit definitions below are expressed by means of the decision policy formulation of Section 2.2:

Definition 3.3 (Subtree errors)**.** *(a) For each stage* $t \in \{1, \dots, T\}$*, we define the* optimal subtree error $\Delta Q^n(x_{..t-1})$ *at node* $n \in \mathcal{N}_t$ *and for a decision policy* $x_{..t-1} \in \Pi_{i=0}^{t-1}(\mathcal{X}_i \cup \widehat{\mathcal{X}}_i)$ *as*

$$\Delta Q^n(x_{..t-1}) = Q_t(x_{..t-1}, x^*_{t..}; \zeta^{\cdot\cdot n}) - \widehat{Q}^n(x_{..t-1}, \widehat{x}^*_{t..}). \tag{34}$$

At the root node, the optimal subtree error *is*

$$\Delta Q^{n_0} = Q_0(x^*_{..T}) - \widehat{Q}^{n_0}(\widehat{x}^*_{..T}). \tag{35}$$

(b) For each stage $t \in \{0, \dots, T\}$*, we define the* suboptimal subtree error $\Delta Q^n_{\text{sub}}(x_{..t})$ *at node* $n \in \mathcal{N}_t$ *and for a decision policy* $x_{..t} \in \Pi_{i=0}^{t}(\mathcal{X}_i \cup \widehat{\mathcal{X}}_i)$ *as*

$$\Delta Q^n_{\text{sub}}(x_{..t}) = \begin{cases} Q_T(x; \zeta^{\cdot\cdot n}) - \widehat{Q}^n(x) & \text{if } t = T; \\ Q_t(x_{..t}, x^*_{t+1..}; \zeta^{\cdot\cdot n}) - \widehat{Q}^n(x_{..t}, \widehat{x}^*_{t+1..}) & \text{if } t \in \{1, \dots, T-1\}; \\ Q_0(x_0, x^*_{1..}) - \widehat{Q}^{n_0}(x_0, \widehat{x}^*_{1..}) & \text{if } t = 0. \end{cases} \tag{36}$$

The subtree errors ΔQ^n_{sub} and ΔQ^n are identically zero for every $n \in \mathcal{N}_T$ (cf. Remark 2.1). In the general setting of the scenario-tree formulation of Section 2.3, we do not know whether the subtree errors have positive or negative values. The node-n_0 optimal subtree error (35) is the optimal-value error that we want to decompose and bound.

The optimal and suboptimal subtree errors at node n gather implicitly all the node errors (optimization and discretization) made at each node $m \in \mathcal{N}(n)$ of the subtree rooted at n. To find an explicit relation between the subtree errors and the node errors, we need to be able to derive a closed-form representation of a quantity at node n from a recursive representation of this quantity over the nodes in the subtree rooted at n. This is the purpose of the following lemma:

Lemma 3.4. *Let a real value* γ^n *be assigned to every node* $n \in \mathcal{N} \setminus \mathcal{N}_T$ *of the scenario tree.*

(a) The sequence $\{\alpha^n : n \in \mathcal{N}\}$ *satisfies the recurrence relation*

$$\alpha^n = \begin{cases} \gamma^n + \displaystyle\sum_{m \in C(n)} w^m \, \alpha^m & \text{if } n \in \mathcal{N} \setminus \mathcal{N}_T; \quad (37) \\ 0 & \text{if } n \in \mathcal{N}_T, \quad (38) \end{cases}$$

if and only if α^n *has a closed-form representation at each node given by*

$$\alpha^n = 0 \quad \forall n \in \mathcal{N}_T \quad \text{and} \quad \alpha^n = \frac{1}{W^n} \sum_{m \in \mathcal{N}(n) \setminus \mathcal{N}_T} W^m \, \gamma^m \quad \forall n \in \mathcal{N} \setminus \mathcal{N}_T, \tag{39}$$

where $\mathcal{N}(n)$ *is the node set of the subtree rooted at* n*.*

(b) If a sequence $\{\beta^n : n \in \mathcal{N}\}$ *satisfies the recurrence relation*

$$\beta^n \leq \begin{cases} \gamma^n + \sum_{m \in C(n)} w^m \, \beta^m & \text{if } n \in \mathcal{N} \setminus \mathcal{N}_T; \quad (40) \\ 0 & \text{if } n \in \mathcal{N}_T, \quad (41) \end{cases}$$

then β^n *has an upper bound at each node* $n \in \mathcal{N} \setminus \mathcal{N}_T$ *given by*

$$\beta^n \leq \frac{1}{W^n} \sum_{m \in \mathcal{N}(n) \setminus \mathcal{N}_T} W^m \, \gamma^m. \quad (42)$$

Parts (a) and (b) will be used in deriving the error decomposition theorem and the error bound theorem, respectively.

Proof. (a) Let $\{u^n\}$ and $\{v^n\}$ denote two sequences satisfying the recurrence relation (37)-(38) and the closed-form (39), respectively. Let us show by backward induction that $u^n = v^n$ holds for every node $n \in \mathcal{N} \setminus \mathcal{N}_T$.

Basis. Take an arbitrary $n \in \mathcal{N}_{T-1}$. We have that $\mathcal{N}(n) \setminus \mathcal{N}_T = \{n\}$, hence it follows from (39) that

$$v^n = \frac{1}{W^n} W^n \, \gamma^n = \gamma^n = u^n. \quad (43)$$

Inductive step. Suppose that $u^m = v^m$ holds for every $m \in \mathcal{N}_t$ for a given $t \in \{1, \ldots, T-1\}$, and take an arbitrary $n \in \mathcal{N}_{t-1}$. Using the following decomposition of $\mathcal{N}(n) \setminus \mathcal{N}_T$:

$$\mathcal{N}(n) \setminus \mathcal{N}_T = \{n\} \cup \Big(\bigcup_{m \in C(n)} \mathcal{N}(m) \setminus \mathcal{N}_T \Big), \quad (44)$$

it follows from (39) that

$$\begin{aligned} v^n &= \gamma^n + \frac{1}{W^n} \sum_{m \in C(n)} \sum_{l \in \mathcal{N}(m) \setminus \mathcal{N}_T} W^l \, \gamma^l & (45) \\ &= \gamma^n + \frac{1}{W^n} \sum_{m \in C(n)} W^m \Big[\frac{1}{W^m} \sum_{l \in \mathcal{N}(m) \setminus \mathcal{N}_T} W^l \, \gamma^l \Big] & (46) \\ &= \gamma^n + \frac{1}{W^n} \sum_{m \in C(n)} W^m \, v^m & (47) \\ &= \gamma^n + \sum_{m \in C(n)} w^m \, u^m & (48) \\ &= u^n, & (49) \end{aligned}$$

where the equality (48) holds by the induction hypothesis and by the relation $W^m = W^n w^m$ for every $m \in C(n)$ (cf. (19)). This proves the inductive step and therefore the final result.

(b) Let $\{\alpha^n\}$ and $\{\beta^n\}$ denote two sequences satisfying the recurrence relation (37)-(38) and (40)-(41), respectively. Let us show by induction that $\beta^n \leq \alpha^n$ holds for every node $n \in \mathcal{N} \setminus \mathcal{N}_T$.

Basis. For every $n \in \mathcal{N}_{T-1}$, it follows from (40)-(41) that $\beta^n \leq \gamma^n = \alpha^n$.

Inductive step. Suppose that $\beta^m \leq \alpha^m$ holds for every node $m \in \mathcal{N}_t$ for a given $t \in \{1, \ldots, T-1\}$, and take an arbitrary $n \in \mathcal{N}_{t-1}$. It follows from (40) and the induction hypothesis that

$$\beta^n \leq \gamma^n + \sum_{m \in C(n)} w^m \, \alpha^m = \alpha^n. \quad (50)$$

This proves the inductive step. The inequality (42) follows immediately using part (a) of this lemma. □

We can now state the main theorem of this section:

Theorem 3.5. *The scenario-tree optimal-value error can be decomposed into a weighted sum of node discretization and optimization errors as follows:*

$$\Delta Q^{n_0} = \sum_{n \in \mathcal{N} \setminus \mathcal{N}_T} W^n \left[\mathrm{E}^n_{\mathrm{opt}}(\widehat{x}^*_{..t(n)-1}) + \mathrm{E}^n_{\mathrm{disc}}(\widehat{x}^*_{..t(n)}) \right], \tag{51}$$

where for $n = n_0$ the term $\mathrm{E}^n_{\mathrm{opt}}(\widehat{x}^*_{..t(n)-1})$ corresponds to $\mathrm{E}^{n_0}_{\mathrm{opt}}$.

Proof. We start by deriving a recurrence relation for the optimal subtree error at node $n \in \mathcal{N} \setminus \mathcal{N}_T$, by considering successively the cases $t = 0$ and $t \in \{1, \ldots, T-1\}$.

At the root node, using successively (23), (31) and (33), we can write ΔQ^{n_0} as follows:

$$\begin{aligned}
\Delta Q^{n_0} &= Q_0(x^*_{..T}) - \widehat{Q}^{n_0}(\widehat{x}^*_{..T}) && (52)\\
&= Q_0(x^*_{..T}) - \sum_{m \in C(n_0)} w^m \, \widehat{Q}^m(\widehat{x}^*_{..T}) && (53)\\
&= Q_0(x^*_{..T}) - Q_0(\widehat{x}^*_0, x^*_{1..}) && (54)\\
&\quad + Q_0(\widehat{x}^*_0, x^*_{1..}) - \sum_{m \in C(n_0)} w^m Q_1(\widehat{x}^*_0, x^*_{1..}; \zeta^m) && (55)\\
&\quad + \sum_{m \in C(n_0)} w^m \left[Q_1(\widehat{x}^*_0, x^*_{1..}; \zeta^m) - \widehat{Q}^m(\widehat{x}^*_{..T}) \right] && (56)\\
&= \mathrm{E}^{n_0}_{\mathrm{opt}} + \mathrm{E}^{n_0}_{\mathrm{disc}}(\widehat{x}^*_0) + \sum_{m \in C(n_0)} w^m \, \Delta Q^m(\widehat{x}^*_0). && (57)
\end{aligned}$$

For every $t \in \{1, \ldots, T-1\}$ and every $n \in \mathcal{N}_t$, using successively (23), (32) and (30), we can write $\Delta Q^n(\widehat{x}^*_{..t-1})$ as follows:

$$\begin{aligned}
\Delta Q^n(\widehat{x}^*_{..t-1}) &= Q_t(\widehat{x}^*_{..t-1}, x^*_{t..}; \zeta^{..n}) - \widehat{Q}^n(\widehat{x}^*_{..T}) && (58)\\
&= Q_t(\widehat{x}^*_{..t-1}, x^*_{t..}; \zeta^{..n}) - \sum_{m \in C(n)} w^m \, \widehat{Q}^m(\widehat{x}^*_{..T}) && (59)\\
&= Q_t(\widehat{x}^*_{..t-1}, x^*_{t..}; \zeta^{..n}) - Q_t(\widehat{x}^*_{..t}, x^*_{t+1..}; \zeta^{..n}) && (60)\\
&\quad + Q_t(\widehat{x}^*_{..t}, x^*_{t+1..}; \zeta^{..n}) - \sum_{m \in C(n)} w^m \, Q_{t+1}(\widehat{x}^*_{..t}, x^*_{t+1..}; \zeta^{..m}) && (61)\\
&\quad + \sum_{m \in C(n)} w^m \Big[Q_{t+1}(\widehat{x}^*_{..t}, x^*_{t+1..}; \zeta^{..m}) - \widehat{Q}^m(\widehat{x}^*_{..T}) \Big] && (62)\\
&= \mathrm{E}^n_{\mathrm{opt}}(\widehat{x}^*_{..t-1}) + \mathrm{E}^n_{\mathrm{disc}}(\widehat{x}^*_{..t}) + \sum_{m \in C(n)} w^m \, \Delta Q^m(\widehat{x}^*_{..t}).
\end{aligned}$$

Finally, by defining

$$\gamma^n = \begin{cases} \mathrm{E}^{n_0}_{\mathrm{opt}} + \mathrm{E}^{n_0}_{\mathrm{disc}}(\widehat{x}^*_0) & \text{if } n = n_0; \quad (63)\\ \mathrm{E}^n_{\mathrm{opt}}(\widehat{x}^*_{..t(n)-1}) + \mathrm{E}^n_{\mathrm{disc}}(\widehat{x}^*_{..t(n)}) & \text{if } n \in \mathcal{N}^* \setminus \mathcal{N}_T, \quad (64) \end{cases}$$

we see that the sequence $\{\Delta Q^n : n \in \mathcal{N}\}$ satisfies the recurrence relation (37)-(38) of Lemma 3.4(a) (recall that $\Delta Q^n = 0$ for every $n \in \mathcal{N}_T$; cf. Remark 2.1). Thus, the decomposition (51) follows directly from (39) applied at the root node. □

4 Node-by-node upper bound on the optimal-value error

The error decomposition of Theorem 3.5, although useful to enlighten the contributions of two types of errors in the optimal-value error, cannot be directly used to guide the generation of scenario trees. The reason is that it features node optimization errors, which are difficult to quantity since they depend on the scenario tree solely via the optimal policy $\widehat{x}^*$. Node discretization errors, conversely, depend directly on the characteristics of a scenario tree, i.e., the tree structure, the discretization points and weights. Moreover, discretization errors enjoy a large literature in numerical integration where they are also referred to as integration errors. In this paper, we use the term "discretization error" when the integrand is the recourse functions (cf. Definition 3.2) and "integration error" when the integrand is any integrable function (cf. Definition 4.4).

The main result of this section is Theorem 4.3, which provides an upper bound on the optimal-value that features only discretization errors. Its derivation does not rely on the decomposition of Theorem 3.5, it is based on the following two lemmas.

Lemma 4.1. *For each stage $t \in \{1, \dots, T\}$, node $n \in \mathcal{N}_t$ and decision policy $x_{..t-1} \in \Pi_{i=0}^{t-1}(\mathcal{X}_i \cup \widehat{\mathcal{X}}_i)$, the following holds:*

$$|\Delta Q^n(x_{..t-1})| \leq \max_{u \in \{\widehat{x}_t^*, x_t^*\}} |\Delta Q^n_{\text{sub}}(x_{..t-1}, u)|, \tag{65}$$

and at the root node:

$$|\Delta Q^{n_0}| \leq \max_{u \in \{\widehat{x}_0^*, x_0^*\}} |\Delta Q^{n_0}_{\text{sub}}(u)|. \tag{66}$$

Proof. If $\Delta Q^{n_0} \geq 0$, then it follows from (17) that

$$\begin{aligned} |\Delta Q^{n_0}| &= Q_0(x^*_{..T}) - \widehat{Q}^{n_0}(\widehat{x}^*_{..T}) && (67)\\ &\leq Q_0(x^*_{..T}) - \widehat{Q}^{n_0}(x_0^*, \widehat{x}^*_{1..}) && (68)\\ &= \Delta Q^{n_0}_{\text{sub}}(x_0^*). && (69) \end{aligned}$$

Conversely, if $\Delta Q^{n_0} < 0$, then it follows from (27) that

$$\begin{aligned} |\Delta Q^{n_0}| &= -Q_0(x^*_{..T}) + \widehat{Q}^{n_0}(\widehat{x}^*_{..T}) && (70)\\ &\leq -Q_0(\widehat{x}_0^*, x^*_{1..}) + \widehat{Q}^{n_0}(\widehat{x}^*_{..T}) && (71)\\ &= -\Delta Q^{n_0}_{\text{sub}}(\widehat{x}_0^*). && (72) \end{aligned}$$

This proves the result at the root node.

Similarly, we show now that the result holds for any $t \in \{1, \dots, T-1\}$, $n \in \mathcal{N}_t$ and $x_{..t-1} \in \Pi_{i=0}^{t}(\mathcal{X}_i \cup \widehat{\mathcal{X}}_i)$. If $\Delta Q^n(x_{..t-1}) \geq 0$, then it follows from (16) that

$$\begin{aligned} |\Delta Q^n(x_{..t-1})| &= Q_t(x_{..t-1}, x^*_{t..}; \zeta^{..n}) - \widehat{Q}^n(x_{..t-1}, \widehat{x}^*_{t..}) && (73)\\ &\leq Q_t(x_{..t-1}, x^*_{t..}; \zeta^{..n}) - \widehat{Q}^n(x_{..t-1}, x_t^*, \widehat{x}^*_{t+1..}) && (74)\\ &= \Delta Q^n_{\text{sub}}(x_{..t-1}, x_t^*). && (75) \end{aligned}$$

If $\Delta Q^n(x_{..t-1}) < 0$, then it follows from (26) that

$$\begin{aligned} |\Delta \widehat{Q}^n(x_{..t-1})| &= -Q_t(x_{..t-1}, x^*_{t..}; \zeta^{..n}) + Q^n(x_{..t-1}, \widehat{x}^*_{t..}) && (76)\\ &\leq -Q_t(x_{..t-1}, \widehat{x}_t^*, x^*_{t+1..}; \zeta^{..n}) + \widehat{Q}^n(x_{..t-1}, \widehat{x}^*_{t..}) && (77)\\ &= -\Delta Q^n_{\text{sub}}(x_{..t-1}, \widehat{x}_t^*). && (78) \end{aligned}$$

This proves the inequality (65) for any $t \in \{1, \dots, T-1\}$. The inequality for $t = T$ holds trivially since $\Delta Q^n(x_{..T-1}) = \Delta Q^n_{\text{sub}}(x_{..T}) = 0$ for all $n \in \mathcal{N}_T$ and $x_{..T} \in \Pi_{i=0}^{T}(\mathcal{X}_i \cup \widehat{\mathcal{X}}_i)$ (cf. Remark 2.1). □

Lemma 4.2. *For each stage $t \in \{1, \ldots, T-1\}$, node $n \in \mathcal{N}_t$ and decision policy $x_{..t-1} \in \Pi_{i=0}^{t-1}(\mathcal{X}_i \cup \widehat{\mathcal{X}}_i)$, the following holds:*

$$|\Delta Q^n(x_{..t-1})| \leq \max_{u \in \{\widehat{x}_t^*, x_t^*\}} |\mathrm{E}^n_{\mathrm{disc}}(x_{..t-1}, u)| + \sum_{m \in C(n)} w^m \max_{u \in \{\widehat{x}_t^*, x_t^*\}} |\Delta Q^m(x_{..t-1}, u)|, \tag{79}$$

and at the root node:

$$|\Delta Q^{n_0}| \leq \max_{u \in \{\widehat{x}_0^*, x_0^*\}} |\mathrm{E}^{n_0}_{\mathrm{disc}}(u)| + \sum_{m \in C(n_0)} w^m \max_{u \in \{\widehat{x}_0^*, x_0^*\}} |\Delta Q^m(u)|. \tag{80}$$

Proof. We first prove the result at the root node. Take an arbitrary $x_0 \in \mathcal{X}_0$ (and recall that $\mathcal{X}_0 = \widehat{\mathcal{X}}_0$). Using successively (23), (34) and (33), we can write $\Delta Q^{n_0}_{\mathrm{sub}}(x_0)$ as

$$\begin{aligned}
\Delta Q^{n_0}_{\mathrm{sub}}(x_0) &= Q_0(x_0, x^*_{1..}) - \widehat{Q}^{n_0}(x_0, \widehat{x}^*_{1..}) && (81)\\
&= Q_0(x_0, x^*_{1..}) - \sum_{m \in C(n_0)} w^m \, \widehat{Q}^m(x_0, \widehat{x}^*_{1..}) && (82)\\
&= Q_0(x_0, x^*_{1..}) - \sum_{m \in C(n_0)} w^m \big[Q_1(x_0, x^*_{1..}; \zeta^m) - \Delta Q^m(x_0)\big] && (83)\\
&= \mathrm{E}^{n_0}_{\mathrm{disc}}(x_0) + \sum_{m \in C(n_0)} w^m \Delta Q^m(x_0). && (84)
\end{aligned}$$

Combining the above equality with the inequality (66), in the particular case for which $x_0 \in \{\widehat{x}_0^*, x_0^*\}$, and applying the triangle inequality yields the result at the root node.

Similarly, we show now that the result holds for any $t \in \{1, \ldots, T-1\}$, $n \in \mathcal{N}_t$ and $x_{..t} \in \Pi_{i=0}^{t}(\mathcal{X}_i \cup \widehat{\mathcal{X}}_i)$. Using successively (23), (34) and (32), we can write $\Delta Q^n_{\mathrm{sub}}(x_{..t})$ as

$$\begin{aligned}
\Delta Q^n_{\mathrm{sub}}(x_{..t}) &= Q_t(x_{..t}, x^*_{t+1..}; \zeta^{\cdot..n}) - \widehat{Q}^n(x_{..t}, \widehat{x}^*_{t+1..}) && (85)\\
&= Q_t(x_{..t}, x^*_{t+1..}; \zeta^{\cdot..n}) - \sum_{m \in C(n)} w^m \, \widehat{Q}^m(x_{..t}, \widehat{x}^*_{t+1..}) && (86)\\
&= Q_t(x_{..t}, x^*_{t+1..}; \zeta^{\cdot..n}) - \sum_{m \in C(n)} w^m \big[Q_{t+1}(x_{..t}, x^*_{t+1..}; \zeta^{\cdot..m}) - \Delta Q^m(x_{..t})\big] && (87)\\
&= \mathrm{E}^n_{\mathrm{disc}}(x_{..t}) + \sum_{m \in C(n)} w^m \Delta Q^m(x_{..t}). && (88)
\end{aligned}$$

From the triangle inequality it follows that

$$|\Delta Q^n_{\mathrm{sub}}(x_{..t})| \leq |\mathrm{E}^n_{\mathrm{disc}}(x_{..t})| + \sum_{m \in C(n)} w^m |\Delta Q^m(x_{..t})|. \tag{89}$$

In the particular case for which $x_{..t} = (x_{..t-1}, u)$, with $u \in \{\widehat{x}_t^*, x_t^*\}$, we combine the above inequality with the inequality (65) to obtain

$$\begin{aligned}
|\Delta Q^n(x_{..t-1})| &\leq \max_{u \in \{\widehat{x}_t^*, x_t^*\}} |\Delta Q^n_{\mathrm{sub}}(x_{..t-1}, u)| && (90)\\
&\leq \max_{u \in \{\widehat{x}_t^*, x_t^*\}} |\mathrm{E}^n_{\mathrm{disc}}(x_{..t-1}, u)| + \sum_{m \in C(n)} w^m \max_{u \in \{\widehat{x}_t^*, x_t^*\}} |\Delta Q^m(x_{..t-1}, u)|. && (91)
\end{aligned}$$

This proves the result for any $t \in \{1, \ldots, T-1\}$. □

Theorem 4.3. *The scenario-tree optimal-value error is bounded by a weighted sum of node discretization errors as follows:*

$$|\Delta Q^{n_0}| \leq \sum_{n \in \mathcal{N} \setminus \mathcal{N}_T} W^n \max_{u \in \Pi_{i=0}^{t(n)} \{\widehat{x}_i^*, x_i^*\}} |\mathrm{E}^n_{\mathrm{disc}}(u)|. \tag{92}$$

Proof. Take an arbitrary $t \in \{1, \ldots, T-1\}$ and $n \in \mathcal{N}_t$. Using the inequality (79) in the particular case for which $x_{..t-1} \in \Pi_{i=0}^{t-1} \{\widehat{x}_i^*, x_i^*\}$ yields

$$\begin{aligned} \max_{v \in \Pi_{i=0}^{t-1} \{\widehat{x}_i^*, x_i^*\}} |\Delta Q^n(v)| \leq & \max_{v \in \Pi_{i=0}^{t-1} \{\widehat{x}_i^*, x_i^*\}} \Big(\max_{u \in \{\widehat{x}_t^*, x_t^*\}} |\mathrm{E}^n_{\mathrm{disc}}(v,u)| \Big) && (93) \\ & + \sum_{m \in C(n)} w^m \max_{v \in \Pi_{i=0}^{t-1} \{\widehat{x}_i^*, x_i^*\}} \Big(\max_{u \in \{\widehat{x}_t^*, x_t^*\}} |\Delta Q^m(v,u)| \Big) \\ = & \max_{(v,u) \in \Pi_{i=0}^{t} \{\widehat{x}_i^*, x_i^*\}} |\mathrm{E}^n_{\mathrm{disc}}(v,u)| && (94) \\ & + \sum_{m \in C(n)} w^m \max_{(v,u) \in \Pi_{i=0}^{t} \{\widehat{x}_i^*, x_i^*\}} |\Delta Q^m(v,u)|. \end{aligned}$$

At all nodes $n \in \mathcal{N}_T$ the following holds trivially (cf. Remark 2.1):

$$\max_{v \in \Pi_{i=0}^{T-1} \{\widehat{x}_i^*, x_i^*\}} |\Delta Q^n(v)| = 0. \tag{95}$$

Finally, by defining

$$\beta^n = \begin{cases} \max\limits_{v \in \Pi_{i=0}^{t(n)-1} \{\widehat{x}_i^*, x_i^*\}} |\Delta Q^n(v)| & \text{if } n \in \mathcal{N}^*; \quad (96) \\ |\Delta Q^{n_0}| & \text{if } n = n_0, \quad (97) \end{cases}$$

and

$$\gamma^n = \max_{w \in \Pi_{i=0}^{t(n)} \{\widehat{x}_i^*, x_i^*\}} |\mathrm{E}^n_{\mathrm{disc}}(w)|, \quad \text{for } n \in \mathcal{N} \setminus \mathcal{N}_T, \tag{98}$$

we see that the sequence $\{\beta^n : n \in \mathcal{N}\}$ satisfies the recurrence relation (40)-(41) of Lemma 3.4(b). Thus, the bound (92) follows directly from (42) applied at the root node. □

Bound in terms of worst-case integration errors

We want now to express the bound (92) as a weighted sum of worst-case integration errors in some function sets. To this end, we first introduce the notion of *integration error* $\mathcal{E}^n(f)$ at node $n \in \mathcal{N} \setminus \mathcal{N}_T$, which represents the error made by using a scenario tree to approximate numerically the conditional expectation of $f(\boldsymbol{\xi}_{t(n)+1})$ given $\boldsymbol{\xi}_{..t(n)} = \zeta^{\cdot\cdot n}$, where f is an appropriately integrable function. The node integration error generalizes the node discretization error of Definition 3.2 to the class of all integrable functions.

In the following, $\mathcal{L}^1(\Xi_1; \mathbb{R})$ denotes the set of all functions $f : \Xi_1 \to \mathbb{R}$ integrable with respect to the distribution of $\boldsymbol{\xi}_1$ and $\mathcal{L}^1(\Xi_{t+1}(\xi_{..t}); \mathbb{R})$ denotes the set of all functions $f : \Xi_{t+1}(\xi_{..t}) \to \mathbb{R}$ integrable with respect to the conditional distribution of $\boldsymbol{\xi}_{t+1}$ given $\boldsymbol{\xi}_{..t} = \xi_{..t}$.

Definition 4.4 (Node integration error)**.** *For every $t \in \{0, \ldots, T-1\}$, we define the* integration error $\mathcal{E}^n(\cdot)$ *at node $n \in \mathcal{N}_t$ as*

$$\mathcal{E}^n(f) = \mathbb{E}[f(\boldsymbol{\xi}_{t+1}) \,|\, \boldsymbol{\xi}_{..t} = \zeta^{\cdot\cdot n}] - \sum_{m \in C(n)} w^m f(\zeta^m), \quad \forall f \in \mathcal{L}^1(\Xi_{t+1}(\zeta^{\cdot\cdot n}); \mathbb{R}), \tag{99}$$

if $t \in \{1, \ldots, T-1\}$, *and at the root node*

$$\mathcal{E}^{n_0}(f) = \mathbb{E}[f(\boldsymbol{\xi}_1)] - \sum_{m \in C(n_0)} w^m f(\zeta^m), \quad \forall f \in \mathcal{L}^1(\Xi_1; \mathbb{R}). \tag{100}$$

The concept of integration error naturally leads to the concept of *worst-case integration error* $\mathrm{E}^n_{\mathrm{wc}}(\mathcal{G})$ which measures the largest integration error at node $n \in \mathcal{N} \setminus \mathcal{N}_T$ for a non-empty function set $\mathcal{G}$:

$$\mathrm{E}^n_{\mathrm{wc}}(\mathcal{G}) := \sup_{f \in \mathcal{G}} |\mathcal{E}^n(f)|. \tag{101}$$

The following function sets of recourse functions are of particular interest to express the bound:

$$\mathcal{Q}^{n_0} = \big\{Q_1(x_0, x^*_{1..}; \cdot) : x_0 \in \{x^*_0, \widehat{x}^*_0\}\big\}, \tag{102}$$

and for every $t \in \{1, \ldots, T-1\}$ and $n \in \mathcal{N}_t$,

$$\mathcal{Q}^n = \big\{Q_{t+1}(x_{..t}, x^*_{t+1..}; \zeta^{\cdot\cdot n}, \cdot) : x_{..t} \in \Pi^t_{i=0}\{\widehat{x}^*_i, x^*_i\}\big\}. \tag{103}$$

Corollary 4.5 below expresses the bound (92) by means of worst-case integration errors.

Corollary 4.5. *Let* $\mathcal{G}^n$, *for every* $n \in \mathcal{N} \setminus \mathcal{N}_T$, *be any function sets satisfying* $\mathcal{Q}^{n_0} \subseteq \mathcal{G}^{n_0} \subseteq \mathcal{L}^1(\Xi_1; \mathbb{R})$ *and* $\mathcal{Q}^n \subseteq \mathcal{G}^n \subseteq \mathcal{L}^1(\Xi_{t(n)+1}(\zeta^{\cdot\cdot n}); \mathbb{R})$. *The scenario-tree optimal-value error is bounded by a weighted sum of worst-case integration errors as follows:*

$$|\Delta Q^{n_0}| \leq \sum_{n \in \mathcal{N} \setminus \mathcal{N}_T} W^n \, \mathrm{E}^n_{\mathrm{wc}}(\mathcal{G}^n). \tag{104}$$

Proof. The worst-case integration error (101) and the node discretization error of Definition 3.2 are linked as follows:

$$\mathrm{E}^n_{\mathrm{wc}}(\mathcal{Q}^n) = \max_{u \in \Pi^{t(n)}_{i=0}\{\widehat{x}^*_i, x^*_i\}} |\mathrm{E}^n_{\mathrm{disc}}(u)|, \quad \forall n \in \mathcal{N} \setminus \mathcal{N}_T. \tag{105}$$

Thus, Theorem 4.3 directly yields the right-hand side of (104) with $\mathcal{Q}^n$ in place of $\mathcal{G}^n$. Moreover, by definition of the worst-case integration error, we have that $\mathcal{Q}^n \subseteq \mathcal{G}^n$ implies that $\mathrm{E}^n_{\mathrm{wc}}(\mathcal{Q}^n) \leq \mathrm{E}^n_{\mathrm{wc}}(\mathcal{G}^n)$ for every $n \in \mathcal{N} \setminus \mathcal{N}_T$, which completes the proof. □

5 Scenario-tree generation

We want now to highlight why the bound (104) carries relevant information about the structure of the problem, namely, the recourse functions, the constraints and the stochastic process, and how it can be used to smartly design scenario trees.

To facilitate an intuitive interpretation of the decomposition (51) and of the bound (92), we examine the special case of a stochastic programming problem for which the constraints of Condition 1 have a unique solution, i.e., Z_0 and $Z_t(\xi_{..t})$ are singletons. This case is equivalent to a numerical integration problem, where the expectation of a function is approximated by a finite sum, since the equation (1) is no longer relevant, as the supremum is trivial, and only equation (2) remains. The latter computes recursively the expectation of $q(z(\boldsymbol{\xi}_{..T}); \boldsymbol{\xi}_{..T})$ where $z(\xi_{..T})$ denotes the only element in $Z_T(\xi_{..T})$. In this setting, the decomposition (51) and the bound (92) are written respectively as

$$\Delta Q^{n_0} = \sum_{n \in \mathcal{N} \setminus \mathcal{N}_T} W^n \, \mathrm{E}^n_{\mathrm{disc}}(x^*_{..t(n)}) \quad \text{and} \quad |\Delta Q^{n_0}| \leq \sum_{n \in \mathcal{N} \setminus \mathcal{N}_T} W^n \, |\mathrm{E}^n_{\mathrm{disc}}(x^*_{..t(n)})|, \tag{106}$$

because the optimization errors (30)-(31) equal zero for all nodes and the set $\Pi_{i=0}^{t}\{\widehat{x}_i^*, x_i^*\}$ over which the maximum is computed in (92) is the singleton $\{x_{..t}^*\}$.

From this we draw the following conclusion for the case of a numerical integration problem solved by scenario trees:

(i) what matters for the control of ΔQ^{n_0} is the control of each node-n integration error $\mathcal{E}^n(f)$ for $f \in \mathcal{Q}^n$ and where $\mathcal{Q}^n$ contains a *unique* function for each $n \in \mathcal{N} \setminus \mathcal{N}_T$.

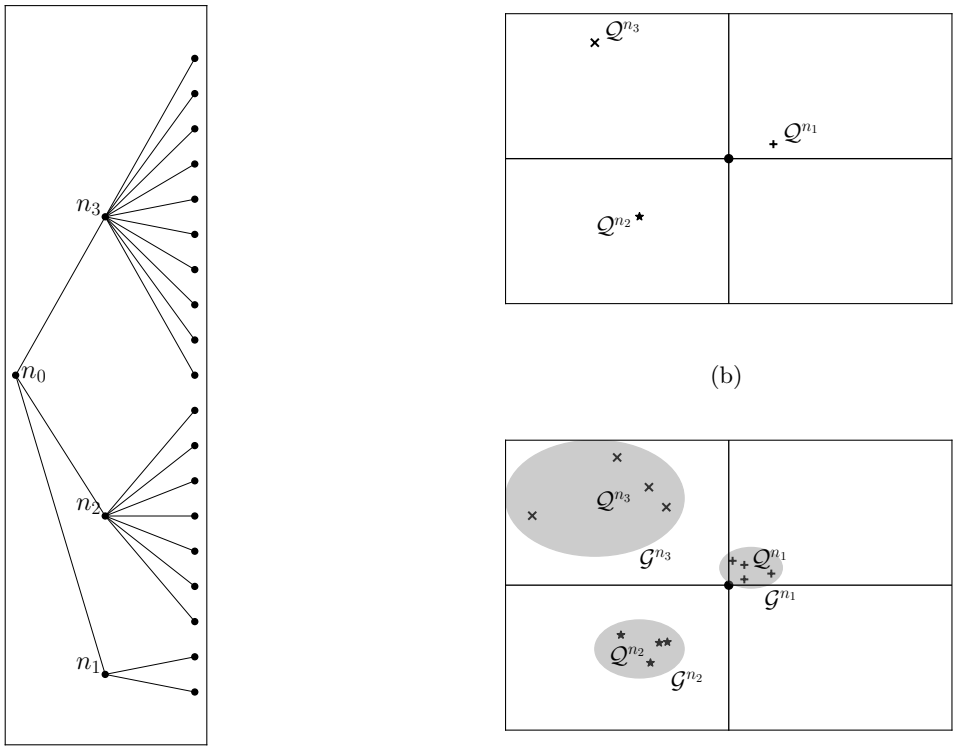
The statement (i) provides insight into the way scenario trees should be built for an efficient numerical integration. Before explaining this, we remind the reader that fundamental results in numerical integration state that the magnitude of the integration error $\mathcal{E}^n(f)$ is linked to the discretization quality of the points and weights as well as to the variability of the integrand f with respect to the probability distribution. This variability is measured, for instance, by the *standard deviation* in the Central Limit Theorem for Monte Carlo sampling or the *Hardy-Krause variation* in the Koksma-Hlawka inequality for quasi-Monte Carlo methods. In any case, an integrand with large variability with respect to the probability distribution typically leads to large integration error, whereas an integrand with low (or no) variability leads to low (or no) integration error. Thus, when building a scenario tree, one should identify nodes n where the variability of the integrand $f \in \mathcal{Q}^n$ is large and assign to them more children nodes $C(n)$ to reduce the integration error, and conversely, assign fewer children nodes $C(m)$ to the nodes m for which $f \in \mathcal{Q}^m$ has low variability. In the limit where $f \in \mathcal{Q}^m$ has no variability, only one node is necessary in $C(m)$ to integrate *exactly* the function numerically. We illustrate qualitatively this in Figure 1(b). The structures of the recourse functions and stochastic process are what determine the variability of each integrand $f \in \mathcal{Q}^n$; indeed, it is the variation of f as a map from $\Xi_{t+1}(\zeta^{\cdot\cdot n})$ to $\mathbb{R}$ as well as the conditional variability of $\boldsymbol{\xi}_{t+1}$ given $\boldsymbol{\xi}_{..t} = \zeta^{\cdot\cdot n}$ that make the expectation in (99) difficult or easy to approximate numerically.

Having established that both the stochastic process and the recourse functions should influence the construction of a scenario tree under the above simplified setting, we consider again the general setting of stochastic programming problems where both equations (1) and (2) need to be approximated by the scenario tree. The difference with the previous setting is the fact that the set $\Pi_{i=0}^{t}\{\widehat{x}_i^*, x_i^*\}$ is no longer reduced to a singleton, hence the point (i) above is now stated as:

(i)' what matters for the control of ΔQ^{n_0} is the control of each node-n integration error $\mathcal{E}^n(f)$ for $f \in \mathcal{Q}^n$ and where $\mathcal{Q}^n$ is a *class* of several functions for each $n \in \mathcal{N} \setminus \mathcal{N}_T$.

Each class $\mathcal{Q}^n$ includes $2^{t(n)+1}$ integrands as it is the cardinality of the set $\Pi_{i=0}^{t(n)}\{\widehat{x}_i^*, x_i^*\}$. Since x^* is not known in practice and $\widehat{x}^*$ depends on the scenario tree, these classes cannot be determined exactly and, for this reason, we have to consider larger classes $\mathcal{G}^n$ that include $\mathcal{Q}^n$ as in Corollary 4.5. Each class $\mathcal{G}^n$ should include $\mathcal{Q}^n$ as tightly as possible to ensure that $\mathrm{E}_{\mathrm{wc}}^n(\mathcal{G}^n)$ is not much larger than $\mathrm{E}_{\mathrm{wc}}^n(\mathcal{Q}^n)$ so that the bound remains tight. How far apart the integrands in $\mathcal{Q}^n$ are is what makes the inclusion tight or loose and this depends on the *structure of the constraints*. Indeed, if the set $Z_{t(n)}(\zeta^{\cdot\cdot n})$ is narrow for some node n, then the decisions obtained with $\widehat{x}^*$ in the realization $\zeta^{\cdot\cdot n}$ cannot fall too far from those obtained with x^*. In this case, the integrands in $\mathcal{Q}^n$ are not far apart and the set $\mathcal{G}^n$ can be chosen so that it includes tightly $\mathcal{Q}^n$. Conversely, if the set $Z_{t(m)}(\zeta^{\cdot\cdot m})$ is large for some other node m, then the elements in $\mathcal{Q}^m$ may be far apart and $\mathcal{G}^m$ can only be chosen in a way that includes loosely $\mathcal{Q}^m$. This is qualitatively illustrated in Figure 1(c), which should be analyzed in comparison with the previous situation of Figure 1(b).

We illustrate in the next example that taking into account the structure of the problem naturally leads to scenario trees with heterogeneous branching suitable to the problem.



(a) (b) (c)

Figure 1: Subfigure (a): Representation of a 3-stage scenario tree with three stage-1 nodes $\mathcal{N}_1 = \{n_1, n_2, n_3\}$ and heterogeneous branching at stage 2. Subfigure (b)-(c): Representation of the integrands in each class $\mathcal{Q}^n$ for $n \in \mathcal{N}_1$ by markers ($\star$, $\times$, $+$) in a plane that represents abstractly a function space of integrable functions. The origin $\bullet$ of the plane is the function with no variability and the distance from the origin to some function f is proportional to the variability of f. (b) describes specifically the setting of *numerical integration problems* for which each $\mathcal{Q}^n$ includes a unique integrand and the node with the most variable integrand (n_3) has the most children nodes whereas the node with the least variable integrand (n_1) has the fewest children nodes. (c) describes the setting of *stochastic programming problems* for which each $\mathcal{Q}^n$ includes 4 integrands and each $\mathcal{G}^n$ (the gray area) should include $\mathcal{Q}^n$ as tightly as possible. Integrands in $\mathcal{Q}^{n_3}$ are far apart hence the inclusion $\mathcal{Q}^{n_1} \subseteq \mathcal{G}^{n_1}$ is loose, whereas integrands in $\mathcal{Q}^{n_1}$ are close to each other so the inclusion $\mathcal{Q}^{n_1} \subseteq \mathcal{G}^{n_1}$ is tight. The most variable integrand is inside the class $\mathcal{G}^{n_3}$ hence n_3 has the most children node, whereas the most variable integrand inside $\mathcal{G}^{n_1}$ is the least variable of the three classes so n_1 is the node with the fewest children nodes.

Example 5.1. Consider the following 3-stage stochastic programming problem:

$$\max_{(y_0,y_1,y_2)\in[0,1]^3}\{-ay_0+\mathbb{E}[-by_1+\boldsymbol{\xi}_2 y_2] : y_1\le y_0,\, y_1\le \boldsymbol{\xi}_1,\, y_2\le y_1\}, \tag{107}$$

where $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$ are random variables (possibly correlated) taking values in $[0,\infty)$ and a, b are positive constant. This problem can be interpreted as follows: at stage 0 a volume y_0 of storage space is reserved at cost a; then at stage 1 a volume y_1 of a commodity is purchased at cost b from a wholesaler and stored in the space reserved previously; the wholesale supply available is random and given by $\boldsymbol{\xi}_1$; finally at stage 2 the commodity is sold to customers at the random price $\boldsymbol{\xi}_2$.

Suppose that the scenario tree contains N stage-1 nodes $\mathcal{N}_1=\{n_1,\dots,n_N\}$ indexed such that $\zeta^{n_1}\le\cdots\le\zeta^{n_N}$. Let us compute the optimal number of children nodes $|C(n_i)|$ for each $i\in\{1,\dots,N\}$ given by minimizing the bound (104).

The stage-2 recourse function $\widetilde{Q}_2^*$ and the set $Z_1(\xi_1)$ of all feasible decisions up to stage 1 are given by

$$\widetilde{Q}_2^*(y_0,y_1;\xi_1,\xi_2)=-ay_0-by_1+\xi_2 y_1, \tag{108}$$

$$Z_1(\xi_1)=\{(y_0,y_1)\in[0,1]^2 : 0\le y_1\le\min(y_0,\xi_1)\}. \tag{109}$$

Since the scenario-tree optimal decisions at nodes (n_0,n_i) necessarily belong to $Z_1(\zeta^{n_i})$, we can define the class of functions $\mathcal{G}^{n_i}$ as

$$\mathcal{G}^{n_i}:=\{\widetilde{Q}_2^*(y_0,y_1;\zeta^{n_i},\cdot) : (y_0,y_1)\in Z_1(\zeta^{n_i})\}, \tag{110}$$

which ensures that $\mathcal{Q}^{n_i}\subseteq\mathcal{G}^{n_i}$ for all $i\in\{1,\dots,N\}$. Defined this way, these classes also satisfy $\mathcal{G}^{n_1}\subseteq\mathcal{G}^{n_2}\subseteq\cdots\subseteq\mathcal{G}^{n_N}$, which in turn implies that $\mathrm{E}_{\mathrm{wc}}^{n_1}(\mathcal{G}^{n_1})\le\mathrm{E}_{\mathrm{wc}}^{n_2}(\mathcal{G}^{n_2})\le\cdots\le\mathrm{E}_{\mathrm{wc}}^{n_N}(\mathcal{G}^{n_N})$, i.e., the structure of the constraints is such that the worst-case integration error at node n_i increases with i. Let us derive the closed-form formula for $\mathrm{E}_{\mathrm{wc}}^{n_i}(\mathcal{G}^{n_i})$. The integration error at node n_i is

$$\mathcal{E}^{n_i}\big(\widetilde{Q}_2^*(y_0,y_1;\zeta^{n_i},\cdot)\big)=\mathbb{E}[\widetilde{Q}_2^*(y_0,y_1;\zeta^{n_i},\boldsymbol{\xi}_2)|\boldsymbol{\xi}_1=\zeta^{n_i}]-\sum_{m\in C(n_i)}w^m\,\widetilde{Q}_2^*(y_0,y_1;\zeta^{n_i},\zeta^m) \tag{111}$$

$$=y_1\Big(\mathbb{E}[\boldsymbol{\xi}_2|\boldsymbol{\xi}_1=\zeta^{n_i}]-\sum_{m\in C(n_i)}w^m\zeta^m\Big), \tag{112}$$

where at the second equality we consider that the weights $\{w^m : m\in C(n_i)\}$ are *normalized*, i.e., $\sum_{m\in C(n_i)}w^m=1$, which allows us to remove the constant terms of $\widetilde{Q}_2^*$. Suppose that we know a discretization method that generates points and weights $\{(\zeta^m,w^m) : m\in C(n_i)\}$ for numerically integrating the functions in $\mathcal{G}^{n_i}$ such that the integration error takes the form:

$$\Big|\mathbb{E}[\boldsymbol{\xi}_2|\boldsymbol{\xi}_1=\zeta^{n_i}]-\sum_{m\in C(n_i)}w^m\zeta^m\Big|=\frac{\mathcal{V}(\boldsymbol{\xi}_2|\zeta^{n_i})}{|C(n_i)|^\alpha}, \tag{113}$$

where $\mathcal{V}(\boldsymbol{\xi}_2|\zeta^{n_i})$ measures the *conditional variability* of $\boldsymbol{\xi}_2$ given ζ^{n_i} (this measure can be the standard deviation, the Hardy-Krause variation, etc., as discussed above) and $\alpha>0$ is the *rate of convergence* of the method (typically $\alpha=1/2$ for Monte Carlo sampling and $\alpha\simeq 1$ in some settings of quasi-Monte Carlo methods; see, e.g., Lemieux [22]). The worst-case integration error $\mathrm{E}_{\mathrm{wc}}^{n_i}(\mathcal{G}^{n_i})$ is given by

$$\mathrm{E}_{\mathrm{wc}}^{n_i}(\mathcal{G}^{n_i})=\sup_{(y_0,y_1)\in Z_1(\zeta^{n_i})}\mathcal{E}^{n_i}\big(\widetilde{Q}_2^*(y_0,y_1;\zeta^{n_i},\cdot)\big)=\frac{\zeta^{n_i}\mathcal{V}(\boldsymbol{\xi}_2|\zeta^{n_i})}{|C(n_i)|^\alpha}. \tag{114}$$

Thus, the optimal number of children nodes $M_i := |C(n_i)|$ for a total number of K scenarios given by minimizing the bound (104) is the optimal solution of

$$\min_{(M_1,\ldots,M_N)\in\mathbb{N}_+^N} \quad \sum_{i=1}^{N} w^{n_i} \frac{\zeta^{n_i}\mathcal{V}(\boldsymbol{\xi}_2|\zeta^{n_i})}{M_i^{\alpha}} \quad \text{subject to} \quad \sum_{i=1}^{N} M_i \leq K. \tag{115}$$

The optimal M_i is therefore determined by the value of the product $w^{n_i}\zeta^{n_i}\mathcal{V}(\boldsymbol{\xi}_2|\zeta^{n_i})$; the larger the product, the more children nodes in $C(n_i)$. If the product is close to zero, then only few nodes are necessary in $C(n_i)$ to keep the integration error $\mathcal{E}^{n_i}(\cdot)$ at almost zero. This will occur if $w^{n_i} \simeq 0$ (i.e., the node n_i is negligible in the scenario tree), or $\zeta^{n_i} \simeq 0$ (i.e., the wholesale supply available at node n_i is close to zero, hence almost no commodity is sold at stage 2), or $\mathcal{V}(\boldsymbol{\xi}_2|\zeta^{n_i}) \simeq 0$ (i.e., the value of the price $\boldsymbol{\xi}_2$ conditional to ζ^{n_i} is almost *not* uncertain). It makes intuitively sense that in any of the three above situations the discretization at n_i should be done with few scenarios.

6 Conclusion

As of today, solving a stochastic programming problem within a given range of error requires a scenario tree of a size that grows fast with the number of stages and the dimension of the random vectors. We believe that this occurs because current scenario-tree generation methods focus mostly on approximating the stochastic process, with little or no regard to the structure of the optimization problem itself, and often restrict their attention to tree structures with regular branching. This paper aims at showing that scenario-tree generation methods could be improved by also taking into account the structure of the problem, namely, the objective function and the constraints, and by tailoring methods to specific classes of problems. The two theorems on the optimal-value error derived in this paper pave the way to designing such methods.

The first theorem is an exact decomposition of the optimal-value error as a weighted sum of discretization and optimization errors made at each node of the scenario tree. It shows that an inappropriate discretization at a node where the recourse function is ill-behaved (e.g., with large variability) can contribute to most of the total optimal-value error. The second theorem is an upper bound on the optimal-value error that features only node discretization errors. It shows that the optimal-value error can be controlled by designing scenario trees suitable for numerically integrating classes of functions determined by the structure of the problem. We have illustrated in Example 5.1 how the upper bound can be effectively used to derive scenario trees suitable to the problem, and we have seen that this naturally leads to tree structures with heterogeneous branching. The branching is denser at nodes where the recourse functions have more variability and sparser at nodes where the variability is lower. Based on this observation, a new scenario-tree generation approach is developed and applied to various types of stochastic programming problems in Keutchayan et al. [18].

References

[1] Fabian Bastin, Cinzia Cirillo, and Philippe L. Toint. Convergence theory for nonconvex stochastic programming with an application to mixed logit. *Mathematical Programming*, 108(2-3): 207–234, 2006.

[2] Richard Bellman. *Dynamic Programming*. Princeton University Press, 1957.

[3] Patrizia Beraldi, Maria Elena Bruni, and Domenico Conforti. Designing robust emergency medical service via stochastic programming. *European Journal of Operational Research*, 158 (1):183–193, 2004.

[4] Marida Bertocchi, Giorgio Consigli, and Michael A. H. Dempster. *Stochastic optimization methods in finance and energy*. Springer, 2011.

[5] John R. Birge and Francois Louveaux. *Introduction to stochastic programming*. Springer Science & Business Media, second edition, 2011.

[6] Matthew Colvin and Christos T. Maravelias. A stochastic programming approach for clinical trial planning in new drug development. *Computers & Chemical Engineering*, 32(11):2626–2642, 2008.

[7] Jitka Dupačová, Nicole Gröwe-Kuska, and Werner Römisch. Scenario reduction in stochastic programming. *Mathematical programming*, 95(3):493–511, 2003.

[8] Martin Dyer and Leen Stougie. Computational complexity of stochastic programming problems. *Mathematical Programming*, 106(3):423–432, 2006.

[9] Nicole Growe-Kuska, Holger Heitsch, and Werner Romisch. Scenario reduction and scenario tree construction for power management problems. In *Power Tech Conference Proceedings, 2003 IEEE Bologna*, volume 3, pages 7–pp. IEEE, 2003.

[10] Grani A. Hanasusanto, Daniel Kuhn, and Wolfram Wiesemann. A comment on "computational complexity of stochastic programming problems". *Mathematical Programming*, 159(1):557–569, 2016.

[11] Holger Heitsch and Werner Römisch. Scenario reduction algorithms in stochastic programming. *Computational optimization and applications*, 24(2-3):187–206, 2003.

[12] Holger Heitsch and Werner Römisch. Scenario tree modeling for multistage stochastic programs. *Mathematical Programming*, 118(2):371–406, 2009.

[13] Holger Heitsch, Werner Römisch, and Cyrille Strugarek. Stability of multistage stochastic programs. *SIAM Journal on Optimization*, 17(2):511–525, 2006.

[14] Tito Homem-de Mello. On rates of convergence for stochastic optimization problems under non-independent and identically distributed sampling. *SIAM Journal on Optimization*, 19(2): 524–551, 2008.

[15] Kjetil Høyland and Stein W. Wallace. Generating scenario trees for multistage decision problems. *Management Science*, 47(2):295–307, 2001.

[16] Kjetil Høyland, Michal Kaut, and Stein W. Wallace. A heuristic for moment-matching scenario generation. *Computational Optimization and Applications*, 24(2-3):169–185, 2003.

[17] Julien Keutchayan, Michel Gendreau, and Antoine Saucier. Quality evaluation of scenario-tree generation methods for solving stochastic programming problems. *Computational Management Science*, 14(3):333–365, 2017.

[18] Julien Keutchayan, David Munger, Michel Gendreau, and Fabian Bastin. A new scenario-tree generation approach for multistage stochastic programming problems based on a demerit criterion. *CIRRELT preprint: CIRRELT-2017-74,* `https://www.cirrelt.ca/DocumentsTravail/CIRRELT-2017-74.pdf`, 2017.

[19] Alan J. King and Stein W. Wallace. *Modeling with stochastic programming*. Springer Science & Business Media, 2012.

[20] Matti Koivu. Variance reduction in sample approximations of stochastic programs. *Mathematical programming*, 103(3):463–485, 2005.

[21] Raimund M. Kovacevic, Georg Ch. Pflug, and Maria Teresa Vespucci. *Handbook of risk management in energy production and trading*. Springer, 2013.

[22] Christiane Lemieux. *Monte Carlo and quasi-Monte Carlo sampling*. Springer New York, 2009.

[23] Hernan Leövey and Werner Römisch. Quasi-Monte Carlo methods for linear two-stage stochastic programming problems. *Mathematical Programming*, 151(1):315–345, 2015.

[24] Wai-Kei Mak, David P. Morton, and R. Kevin Wood. Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24(1):47–56, 1999.

[25] Paul Olsen. Discretizations of multistage stochastic programming problems. In Roger J-B Wets, editor, *Stochastic Systems: Modeling, Identification and Optimization, II*, pages 111–124. Springer, 1976.

[26] Teemu Pennanen. Epi-convergent discretizations of multistage stochastic programs. *Mathematics of Operations Research*, 30(1):245–256, 2005.

[27] Teemu Pennanen. Epi-convergent discretizations of multistage stochastic programs via integration quadratures. *Mathematical Programming*, 116(1-2):461–479, 2009.

[28] Georg Ch. Pflug. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical programming*, 89(2):251–271, 2001.

[29] Georg Ch. Pflug and Alois Pichler. Approximations for probability distributions and stochastic optimization problems. In Marida Bertocchi, Giorgio Consigli, and Michael AH Dempster, editors, *Stochastic optimization methods in finance and energy*, pages 343–387. Springer, 2011.

[30] Georg Ch. Pflug and Alois Pichler. A distance for multistage stochastic optimization models. *SIAM Journal on Optimization*, 22(1):1–23, 2012.

[31] Georg Ch. Pflug and Alois Pichler. Dynamic generation of scenario trees. *Computational Optimization and Applications*, 62(3):641–668, 2015.

[32] Warren B. Powell and Huseyin Topaloglu. Stochastic programming in transportation and logistics. In Andrzej Ruszczyński and Alexander Shapiro, editors, *Handbooks in operations research and management science: Stochastic Programming*, volume 10, pages 555–635. Elsevier, 2003.

[33] András Prékopa. *Stochastic programming*. Kluwer Academic Publishers, Dordrecht, Boston, 1995.

[34] R. Tyrrell Rockafellar and Roger J-B. Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[35] Werner Römisch. Stability of stochastic programming problems. In Andrzej Ruszczyński and Alexander Shapiro, editors, *Handbooks in operations research and management science*, volume 10, pages 483–554. Elsevier, 2003.

[36] Andrzej Ruszczyński and Alexander Shapiro, editors. *Stochastic Programming, volume 10 of Handbooks in Operations Research and Management Science*. Elsevier, 2003.

[37] Alexander Shapiro. Monte Carlo sampling methods. In Andrzej Ruszczyński and Alexander Shapiro, editors, *Handbooks in operations research and management science*, volume 10, pages 353–425. Elsevier, 2003.

[38] Alexander Shapiro and Tito Homem-de Mello. A simulation-based approach to two-stage stochastic programming with recourse. *Mathematical Programming*, 81(3):301–325, 1998.

[39] Alexander Shapiro and Tito Homem-de Mello. On the rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs. *SIAM journal on optimization*, 11(1): 70–86, 2000.

[40] Joyce W. Yen and John R. Birge. A stochastic programming approach to the airline crew scheduling problem. *Transportation Science*, 40(1):3–14, 2006.

[41] Li-Yong Yu, Xiao-Dong Ji, and Shou-Yang Wang. Stochastic programming models in financial optimization: A survey. *Advanced Modeling and Optimization*, 5(1):1–26, 2003.

[42] William T. Ziemba. *The stochastic programming approach to asset, liability, and wealth management*. AIMR, Charlottesville, VA, 2003.