Chapter 14
Scenario Tree Generation for Multi-stage Stochastic Programs

Holger Heitsch and Werner Römisch

Abstract We broaden the theoretical basis for generating scenario trees in multi-stage stochastic programming based on stability analysis. Numerical experience for constructing trees of demand and price scenarios in electricity portfolio management of a municipal power utility is also provided.

14.1 Introduction

Many solution methods for stochastic programming models in finance and energy rely on approximating the underlying probability distribution P by a probability measure based on a finite number of scenarios (or atoms). In case of multi-stage models scenarios have to satisfy certain constraints due to the nonanticipativity of decisions. Such constraints lead to tree-structured scenarios. Hence, designing approximation schemes for multi-stage models requires the *generation* of *scenarios* as well as of *trees*.

There is a variety of methods for generating scenarios (see also the introductory overview (Römisch 2003) and Chapter 15, this volume), namely,

– Monte Carlo methods (applied to statistical models or data) (Shapiro 2003b),
– Quasi-Monte Carlo methods (see Lemieux (2009) for a recent exposition),
– optimal quantization methods for probability measures (see Graf and Luschgy (2000)),
– quadrature rules using sparse grids (e.g. Chen and Mehrotra (2008))

based on the underlying probability distribution P. In general, however, the generated scenarios will not exhibit tree structure.

Presently, there exist several approaches for generating such scenario trees. In a number of cases the tree structure is (partially) predetermined and scenarios

W. Römisch (✉)

Institute of Mathematics, Humboldt University Berlin, 10099 Berlin, Germany
e-mail: romisch@math.hu-berlin.de

are generated sequentially by *(conditional) random sampling* (Dempster 2004; Kouwenberg 2001; Shapiro 2003a) or *quasi-Monte Carlo sampling* (Pennanen 2009). Alternatively, scenarios are generated by using bound-based constructions (Frauendorfer 1996; Kuhn 2005), using optimization methods such that certain moments are matched (Høyland and Wallace 2001; Gülpinar et al. 2004) or such that a best approximation of P is obtained in terms of a suitable probability metric (Pflug 2001; Hochreiter and Pflug 2007; Mirkov and Pflug 2007). We also refer to Dupačová et al. (2000) and the references therein for an overview on scenario tree generation.

The approach to scenario tree generation presented in the following does not require restrictive conditions on P. It starts with a number of scenarios generated by one of the methods mentioned above. The branching structure of the tree is not pre-determined, but automatically detected by the algorithms such that a good approximation of P measured in terms of the closeness of optimal values is obtained. The whole approach is based on a quantitative stability analysis of (linear) multi-stage stochastic programs (see Heitsch et al. (2006) and Heitsch and Römisch (2010)). The algorithms rely on applying scenario reduction sequentially over time and are first analyzed in Heitsch and Römisch (2009a). In this chapter we review parts of our earlier work in Sections 14.2 and 14.4, develop new convergence results in Section 14.3, and report on an application to the electricity portfolio management of a municipal power company in Section 14.5. Readers which are mainly interested in algorithms and their application may skip Sections 14.2 and 14.3.

To state the mathematical optimization model, let the periods of the time horizon be denoted $t = 1, \ldots, T$ and $\xi = (\xi_t)_{t=1}^{T}$ be an $\mathbb{R}^d$-valued stochastic process on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The filtration $(\mathcal{F}_t(\xi))_{t=1}^{T}$ associated with ξ is defined by $\mathcal{F}_t(\xi) := \sigma(\xi^t)$ with $\xi^t = (\xi_1, \ldots, \xi_t)$ for $t = 1, \ldots, T$. We assume that $\mathcal{F}_1 = \{\emptyset, \Omega\}$ and $\xi_t \in L_r(\Omega, \mathcal{F}, \mathbb{P})$, i.e., $\mathbb{E}(|\xi_t|^r) < +\infty$ for every $t = 1, \ldots, T$ and some $r \geq 1$. By $P = \mathbb{P}\xi^{-1}$ we denote the probability distribution of ξ and by P_t its tth marginal probability distribution for $t = 1, \ldots, T$, i.e.,

$$P_t(B_t) = \mathbb{P}\xi^{-1}(\varXi_1 \times \cdots \times \varXi_{t-1} \times B_t \times \varXi_{t+1} \times \cdots \times \varXi_T) \quad (B_t \in \mathcal{B}(\varXi_t)),$$

where $\varXi_t \subseteq \mathbb{R}^d$ denotes the support of ξ_t and $\mathcal{B}(\varXi_t)$ the σ-field of its Borel subsets. In particular, $\varXi_1$ denotes the singleton $\varXi_1 = \{\xi_1\}$.

A linear multi-stage stochastic program can be written as

$$\min \left\{ \mathbb{E}\left(\sum_{t=1}^{T} \langle b_t(\xi_t), x_t \rangle \right) \middle| \begin{array}{l} x_t = x_t(\xi_1, \ldots, \xi_t), \ x_t \in X_t, \\ A_{t,0}x_t + A_{t,1}(\xi_t) = h_t(\xi_t) \end{array} (t = 1, \ldots, T) \right\}, \quad (14.1)$$

where x_t is an $\mathbb{R}^{m_t}$-valued decision vector for time period t. The latter is a Borel measurable function depending (only) on $(\xi_1, \ldots, \xi_t) \in \times_{\tau=1}^{t} \varXi_\tau = \varXi^t$, i.e., it depends on the data observed until time t (*nonanticipativity*). In particular, the components of x_1 are *here and now* decisions since x_1 may only depend on ξ_1 which was assumed to be deterministic. The decisions are subject to constraints: each x_t has to be chosen within a given polyhedral set X_t. Moreover, there are dynamic constraints

involving matrices $A_{t,\tau}$, $\tau \in \{0, 1\}$, and right-hand sides h_t. The matrices $A_{t,1}(\cdot)$, the cost coefficients $b_t(\cdot)$, and right-hand sides $h_t(\cdot)$ may depend on ξ_t in an affinely linear way. $\mathbb{E}$ denotes expectation with respect to $\mathbb{P}$, i.e., the objective corresponds to the expected total costs of a decision vector $(x_1, \ldots, x_T)$.

14.2 Stability of Multi-stage Stochastic Programs

Studying stability of the multi-stage stochastic program (14.1) consists in regarding it as an optimization problem in the infinite dimensional linear space $\times_{t=1}^{T} L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{m_t})$. This is a Banach space when endowed with the norm

$$\|x\|_{r'} := \left(\sum_{t=1}^{T} \mathbb{E}\left[|x_t|^{r'} \right] \right)^{1/r'} \quad \text{for } r' \in [1, +\infty),$$

$$\|x\|_{\infty} := \max_{t=1,\ldots,T} \text{ess sup } |x_t|,$$

where $|.|$ denotes some norm on the relevant Euclidean spaces and ess sup $|x_t|$ denotes the essential supremum of $|x_t|$, i.e., the smallest constant C such that $|x_t| \leq C$ holds $\mathbb{P}$-almost surely. Analogously, ξ can be understood as an element of the Banach space $\times_{t=1}^{T} L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$ with norm $\|\xi\|_r$. For the integrability numbers r and r' it will be imposed that

$$r := \begin{cases} \in [1, +\infty) & \text{, if only costs or only right-hand sides are random} \\ 2 & \text{, if only costs and right-hand sides are random} \\ T & \text{, if all technology matrices are random} \end{cases},$$

$$(14.2)$$

$$r' := \begin{cases} \frac{r}{r-1} & \text{, if only costs are random} \\ r & \text{, if only right-hand sides are random} \\ +\infty & \text{, if all technology matrices are random} \end{cases},$$

with regard to problem (14.1). The choice of r and the definition of r' are motivated by the knowledge of existing moments of the input process ξ, by having the stochastic program well defined (in particular, such that $\langle b_t(\xi_t), x_t \rangle$ is integrable for every decision x_t and $t = 1, ..., T$), and by satisfying the conditions (A2) and (A3) (see below).

Since r' depends on r and our assumptions will depend on both r and r', we will add some comments on the choice of r and its interplay with the structure of the underlying stochastic programming model. To have the stochastic program well defined, the existence of certain moments of ξ has to be required. This fact is well known for the two-stage situation (see, e.g., Ruszczyński and Shapiro (2003), chapter 2). If either right-hand sides or costs in a multi-stage model (14.1) are random, it is sufficient to require $r \geq 1$. The flexibility in case that the stochastic process ξ has moments of order $r > 1$ may be used to choose r' as small as possible in order to weaken condition (A3) (see below) on the feasible set. If the linear

stochastic program is fully random (i.e., costs, right-hand sides, and technology matrices are random), one needs $r \geq T$ to have the model well defined and no flexibility w.r.t. r' remains.

Next we introduce some notation. We set $s := Td$ and $m := \sum_{t=1}^{T} m_t$. Let

$$F(\xi, x) := \mathbb{E}\left[\sum_{t=1}^{T} \langle b_t(\xi_t), x_t \rangle\right]$$

denote the objective function defined on $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s) \times L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ and let

$$\mathcal{X}(\xi) := \left\{ x \in \times_{t=1}^{T} L_{r'}(\Omega, \sigma(\xi^t), \mathbb{P}; \mathbb{R}^{m_t}) \mid x_t \in \mathcal{X}_t(x_{t-1}; \xi_t) \text{ a.s. } (t = 1, ..., T) \right\}$$

denote the set of feasible elements of (14.1) with $x_0 \equiv 0$ and

$$\mathcal{X}_t(x_{t-1}; \xi_t) := \left\{ x_t \in \mathbb{R}^{m_t} : x_t \in X_t, A_{t,0}x_t + A_{t,1}(\xi_t)x_{t-1} = h_t(\xi_t) \right\}$$

denoting the tth feasibility set for every $t = 1, ..., T$. This allows to rewrite the stochastic program (14.1) in the short form

$$\min \{ F(\xi, x) : x \in \mathcal{X}(\xi) \}. \tag{14.3}$$

In the following, we need the optimal value

$$v(\xi) = \inf \{ F(\xi, x) : x \in \mathcal{X}(\xi) \}$$

for every $\xi \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ and, for any $\varepsilon \geq 0$, the *ε-approximate solution set* (level set)

$$S_\varepsilon(\xi) := \{ x \in \mathcal{X}(\xi) : F(\xi, x) \leq v(\xi) + \varepsilon \}$$

of the stochastic program (14.3). Since, for $\varepsilon = 0$, the set $S_\varepsilon(\xi)$ coincides with the set solutions to (14.3), we will also use the notation $S(\xi) := S_0(\xi)$. The following conditions will be imposed on (14.3):

(A1) The numbers r, r' are chosen according to (14.2) and $\xi \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$.

(A2) There exists a $\delta > 0$ such that for any $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ satisfying $\|\tilde{\xi} - \xi\|_r \leq \delta$, any $t = 2, ..., T$ and any $x_\tau \in L_{r'}(\Omega, \sigma(\tilde{\xi}^\tau), \mathbb{P}; \mathbb{R}^{m_\tau})$ ($\tau = 1, ..., t-1$) satisfying $x_\tau \in \mathcal{X}_\tau(x_{\tau-1}; \tilde{\xi}_\tau)$ a.s. (where $x_0 = 0$), there exists $x_t \in L_{r'}(\Omega, \sigma(\tilde{\xi}^t), \mathbb{P}; \mathbb{R}^{m_t})$ satisfying $x_t \in \mathcal{X}_t(x_{t-1}; \tilde{\xi}_t)$ a.s. (*relatively complete recourse locally around* ξ).

(A3) The optimal values $v(\tilde{\xi})$ of (14.3) with input $\tilde{\xi}$ are finite for all $\tilde{\xi}$ in a neighborhood of ξ and the objective function F is *level-bounded locally uniformly at* ξ, i.e., for some $\varepsilon_0 > 0$ there exists a $\delta > 0$ and a bounded

subset B of $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ such that $S_{\varepsilon_0}(\tilde{\xi})$ is contained in B for all $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ with $\|\tilde{\xi} - \xi\|_r \leq \delta$.

For any $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ sufficiently close to ξ in L_r, condition (A2) implies the existence of some feasible $\tilde{x}$ in $\mathcal{X}(\tilde{\xi})$ and (14.2) implies the finiteness of the objective $F(\tilde{\xi}, \cdot)$ at any feasible $\tilde{x}$. A sufficient condition for (A2) to hold is the *complete recourse condition* on every recourse matrix $A_{t,0}$, i.e., $A_{t,0}X_t = \mathbb{R}^{n_t}$, $t = 1, ..., T$. The locally uniform level-boundedness of the objective function F is quite standard in perturbation results for optimization problems (see, e.g., Rockafellar and Wets (1998), theorem 1.17). The finiteness condition on the optimal value $v(\xi)$ is not implied by the level boundedness of F for all relevant pairs (r, r'). In general, conditions (A2) and (A3) get weaker for increasing r and decreasing r', respectively.

The first stability result for multi-stage stochastic programs represents a quantitative continuity property of the optimal values. Its main observation is that multi-stage models behave stable at some stochastic input process ξ if both its probability distribution and its filtration are approximated with respect to the L_r-distance and the filtration distance

$$D_{\mathrm{f}}(\xi, \tilde{\xi}) := \sup_{\varepsilon > 0} \inf_{\substack{x \in S_\varepsilon(\xi) \\ \tilde{x} \in S_\varepsilon(\tilde{\xi})}} \sum_{t=2}^{T-1} \max\left\{ \left\| x_t - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})] \right\|_{r'}, \left\| \tilde{x}_t - \mathbb{E}[\tilde{x}_t | \mathcal{F}_t(\xi)] \right\|_{r'} \right\},$$
(14.4)

where $\mathbb{E}[\,\cdot\,|\mathcal{F}_t(\xi)]$ and $\mathbb{E}[\,\cdot\,|\mathcal{F}_t(\tilde{\xi})]$ $(t = 1, ..., T)$ are the corresponding conditional expectations, respectively. Note that for the supremum in (14.4) only small ε's are relevant and that the approximate solution sets are bounded for $\varepsilon \in (0, \varepsilon_0]$ according to (A3).

The following stability result for optimal values of program (14.3) is taken from Heitsch et al. (2006), theorem 2.1.

Theorem 1 *Let (A1), (A2), and (A3) be satisfied and the set X_1 be nonempty and bounded. Then there exist positive constants L and δ such that the estimate*

$$\left| v(\xi) - v(\tilde{\xi}) \right| \leq L \left(\|\xi - \tilde{\xi}\|_r + D_{\mathrm{f}}(\xi, \tilde{\xi}) \right)$$
(14.5)

holds for all random elements $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ with $\|\tilde{\xi} - \xi\|_r \leq \delta$.

The result states that the changes of optimal values are at most proportional to the errors in terms of L_r- and filtration distance when approximating ξ. The corresponding constant L depends on $\|\xi\|_r$ (i.e. the rth moment of ξ), but is not known in general.

The filtration distance has a simpler representation if the approximation $\tilde{\xi}$ of ξ is ξ-*adapted*, i.e., if $\mathcal{F}_t(\tilde{\xi}) \subseteq \mathcal{F}_t(\xi)$ holds for every $t = 1, \ldots, T$. The latter is equivalent to the existence of measurable functions $f_t : \Xi^t \to \Xi^t$ such that

$$\tilde{\xi}^t = f_t(\xi^t) \quad (t = 1, \ldots, T).$$

For ξ-adapted approximations $\tilde{\xi}$ we have

$$D_f(\xi, \tilde{\xi}) = \sup_{\varepsilon > 0} \inf_{x \in S_\varepsilon(\xi)} \sum_{t=2}^{T-1} \left\| x_t - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})] \right\|_{r'} \tag{14.6}$$

and if, in addition, the solution set $S(\xi)$ is nonempty, we obtain

$$D_f(\xi, \tilde{\xi}) = \inf_{x \in S(\xi)} \sum_{t=2}^{T-1} \left\| x_t - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})] \right\|_{r'}. \tag{14.7}$$

The latter representation allows the conclusion

$$D_f(\xi, \tilde{\xi}) \leq \sum_{t=2}^{T-1} \left\| x_t - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})] \right\|_{r'} \tag{14.8}$$

for any solution $x = (x_1, x_2, \ldots, x_T)$ of (14.3). For a given solution x of (14.3) there exist Borel measurable functions $g_t : \Xi^t \to \mathbb{R}^{m_t}$ such that

$$\mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})] = \mathbb{E}[x_t | \tilde{\xi}^t] = g_t(\tilde{\xi}^t) \quad (t = 1, \ldots, T),$$

where $g_1(\tilde{\xi}_1) = g_1(\xi_1) = x_1$ and $g_t(\xi^t) = x_t$. In general, further properties of the functions

$$g_t(y_1, \ldots, y_t) = \mathbb{E}[x_t | \tilde{\xi}^t = (y_1, \ldots, y_t)] = \mathbb{E}[x_t | f_t(\xi^t) = (y_1, \ldots, y_t)],$$

i.e., of the *conditional expectations of x_t under the condition that $\tilde{\xi}^t$ equals* $(y_1, \ldots, y_t)$, are not known. Since x_t is a (measurable) function of ξ^t, the function value $g_t(y_1, \ldots, y_t)$ may be computed via the (multivariate) probability distribution P of ξ.

Unfortunately, in general, there is no solution $x \in S(\xi)$ such that the functions x_t depend continuously on ξ^t for every $t = 1, \ldots, T$ (cf. the discussion in Rockafellar and Wets (1974)). Sometimes, however, the functional dependence of x_t on ξ^t is of a specific form as in the following situation (Garstka and Wets 1974, theorem 4.3).

Proposition 1 *Assume that only right-hand sides in (14.1) are random and that $S(\xi)$ is nonempty. Then there exists $x = (x_1, \ldots, x_T)$ in $S(\xi)$ such that $x_t = \varphi_t(h_1(\xi_1), \ldots, h_t(\xi_t))$ and φ_t is a continuous, piecewise affine function for every $t = 1, \ldots, T$. In particular, x_t is Lipschitz continuous as a function of ξ^t for every $t = 1, \ldots, T$.*

This motivates the following condition on the conditional distributions of ξ and on the ξ-adapted approximation $\tilde{\xi}$ of ξ.

(A4) For each $t \in \{1, \ldots, T\}$ and each pair (Φ_t, f_t) of Lipschitz continuous functions $\Phi_t : \Xi^t \to \mathbb{R}^{m_t}$ and $f_t : \Xi^t \to \Xi^t$, the function

$$g_t(y_1, \ldots, y_t) = \mathbb{E}[\Phi_t(\xi^t)| f_t(\xi^t) = (y_1, \ldots, y_t)] \quad (14.9)$$

is Lipschitz continuous on Ξ^t.

Then the following result is an immediate consequence of Theorem 1 and Proposition 1 if (A4) is imposed in addition.

Corollary 1 *Let (A1), (A2), (A3), and (A4) be satisfied and X_1 be nonempty and bounded. Assume that only right-hand sides in (14.1) are random and that $S(\xi) \neq \emptyset$. Then there exist positive constants $\hat{L}$ and δ such that*

$$\left| v(\xi) - v(\tilde{\xi}) \right| \leq \hat{L} \|\xi - \tilde{\xi}\|_r \quad (14.10)$$

whenever $\|\xi - \tilde{\xi}\|_r < \delta$ for any ξ-adapted $\tilde{\xi}$ such that $\tilde{\xi}^t = f_t(\xi^t)$, $t = 1, \ldots, T$, with Lipschitz continuous functions $f_t : \Xi^t \to \mathbb{R}^{td}$.

Proof According to Proposition 1 there exists a solution $x = (x_1, \ldots, x_T) \in S(\xi)$ such that $x_t = \Phi_t(\xi^t)$ is a Lipschitz continuous function of ξ^t for every $t = 1, \ldots, T$. Let $\tilde{\xi}$ be ξ-adapted such that $\tilde{\xi}^t = f_t(\xi^t)$, $t = 1, \ldots, T$, where the functions $f_t : \Xi^t \to \Xi^t$ are Lipschitz continuous on Ξ^t. According to (A4) the functions g_t from Ξ^t to $\mathbb{R}^{m_t}$, $t = 1, \ldots, T$, given by (14.9) are Lipschitz continuous. Hence, there exist constants $K_t > 0$ such that

$$|g_t(y_1, \ldots, y_t) - g_t(\tilde{y}_1, \ldots, \tilde{y}_t)| \leq K_t \sum_{\tau=1}^{t} |y_\tau - \tilde{y}_\tau| \quad ((y_1, \ldots, y_t), (\tilde{y}_1, \ldots, \tilde{y}_t) \in \Xi^t)$$

and, hence,

$$D_f(\xi, \tilde{\xi}) \leq \sum_{t=2}^{T-1} \left\| g_t(\xi^t) - g_t(\tilde{\xi}^t) \right\|_r \leq \sum_{t=2}^{T-1} K_t \sum_{\tau=1}^{t} \mathbb{E}(|\xi_\tau - \tilde{\xi}_\tau|^r)^{\frac{1}{r}} \leq K \|\xi - \tilde{\xi}\|_r$$

for some suitably large constant K. Together with Theorem 1 we obtain (14.10) with $\hat{L} = LK$. $\qquad \square$

We note that our condition (A4) is similar to assumption 2.6 in Küchler (2008) and Corollary 1 reminds of Küchler (2008), theorem 3. We also note that in case of finite supports Ξ_t, $t = 1, \ldots, T$, the functions g_t are necessarily Lipschitz continuous with possibly large Lipschitz constants K_t, $t = 1, \ldots, T$, leading to a large constant $\hat{L}$ in Corollary 1.

Stability results of approximate solutions to (14.3) require a stronger version of the filtration distance D_f, namely,

$$D_{\mathrm{f}}^*(\xi, \tilde{\xi}) := \sup_{x \in \mathcal{B}_\infty} \sum_{t=2}^{T} \left\| \mathbb{E}[x_t | \mathcal{F}_t(\xi)] - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})] \right\|_{r'}, \qquad (14.11)$$

where $\mathcal{B}_\infty := \{x : \Omega \to \mathbb{R}^m : x \text{ is } \mathcal{F}\text{-measurable}, |x(\omega)| \leq 1, \mathbb{P}\text{-almost surely}\}$. Notice that the sum is extended by the additional summand for $t = T$ and that the former infimum is replaced by a supremum with respect to a sufficiently large bounded set. If we require, in addition to assumption (A3), that for some $\varepsilon_0 > 0$ there exist constants $\delta > 0$ and $C > 0$ such that $|\tilde{x}(\omega)| \leq C$ for $\mathbb{P}$-almost every $\omega \in \Omega$ and all $\tilde{x} \in S_{\varepsilon_0}(\tilde{\xi})$ with $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ and $\|\tilde{\xi} - \xi\|_r \leq \delta$, we have

$$D_{\mathrm{f}}(\xi, \tilde{\xi}) \leq C \, D_{\mathrm{f}}^*(\xi, \tilde{\xi}). \qquad (14.12)$$

Sometimes it is sufficient to consider the unit ball in $L_{r'}$ rather than the corresponding ball $\mathcal{B}_\infty$ in L_∞ (cf. Heitsch and Römisch (2009a, 2010)). In contrast to D_{f} the distance D_{f}^* is a metric as it satisfies the triangle inequality.

Next we state a second stability result that represents a calmness property of approximate solution sets (Heitsch and Römisch (2010), theorem 2.4).

Theorem 2 *Let (A1), (A2), and (A3) be satisfied with $r' \in [1, +\infty)$ and the set X_1 be nonempty and bounded. Assume that $S(\xi) \neq \emptyset$. Then there exist $\bar{L} > 0$ and $\bar{\varepsilon} > 0$ such that*

$$d\!l_\infty\left(S_\varepsilon(\xi), S_\varepsilon(\tilde{\xi})\right) \leq \frac{\bar{L}}{\varepsilon}\left(\|\xi - \tilde{\xi}\|_r + D_{\mathrm{f}}^*(\xi, \tilde{\xi})\right) \qquad (14.13)$$

holds for every $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ with $\|\xi - \tilde{\xi}\|_r \leq \delta$ (with $\delta > 0$ from (A3)) and $S(\tilde{\xi}) \neq \emptyset$, and for any $\varepsilon \in (0, \bar{\varepsilon})$. Here, $d\!l_\infty$ denotes the Pompeiu–Hausdorff distance of closed bounded subsets of $L_{r'} = L_{r'}(\Omega.\mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ given by

$$d\!l_\infty(B, \tilde{B}) = \sup_{x \in L_{r'}} \left| d_B(x) - d_{\tilde{B}}(x) \right|$$

with $d_B(x)$ denoting the distance of x to B, i.e., $d_B(x) = \inf_{y \in B} \|x - y\|_{r'}$.

The most restrictive assumption in Theorem 2 is the existence of solutions to both problems. Notice that solutions always exist if the underlying random vector ξ has a finite number of scenarios or if $r' \in (1, +\infty)$. For a more thorough discussion we refer to Heitsch and Römisch (2010), section 2. Notice that the constant $\frac{\bar{L}}{\varepsilon}$ gets larger for decreasing ε and that, indeed, Theorem 2 does not remain true for the Pompeiu–Hausdorff distance of solution sets $S(\xi) = S_0(\xi)$ and $S(\tilde{\xi}) = S_0(\tilde{\xi})$, respectively.

14.3 Tree Approximation Framework and Convergence

We present a general framework for tree approximations of multi-stage stochastic programs in case that empirical estimates of the underlying probability distribution are available and prove convergence using the stability results of Section 14.2.

First, we show that sequences $(\xi^{(n)})$ of ξ-adapted random variables converging to ξ in L_r also converge to ξ in terms of the filtration distance D_f. Recall that $\xi^{(n)}$ is ξ-adapted if $\mathcal{F}_t(\xi^{(n)}) \subseteq \mathcal{F}_t(\xi)$ holds for every $t = 1, \ldots, T$.

Proposition 2 *Let (A1), (A2), and (A3) be satisfied, $r' \in [1, +\infty)$ and $S(\xi)$ be nonempty. Let $(\xi^{(n)})$ be a sequence of ξ-adapted random variables converging to ξ in L_r such that the σ-fields $\mathcal{F}_t(\xi^{(n)})$ are nondecreasing with respect to n for every $t = 2, \ldots, T$. Then it holds*

$$\lim_{n \to \infty} D_f(\xi, \xi^{(n)}) = \lim_{n \to \infty} \inf_{x \in S(\xi)} \sum_{t=2}^{T-1} \left\| x_t - \mathbb{E}[x_t | \mathcal{F}_t(\xi^{(n)})] \right\|_{r'} = 0.$$

Proof By $\hat{\mathcal{F}}_t$ we denote the smallest σ-field containing $\mathcal{F}_t(\xi^{(n)})$ for every $n \in \mathbb{N}$, i.e.,

$$\hat{\mathcal{F}}_t := \sigma \left(\bigcup_{n \in \mathbb{N}} \mathcal{F}_t(\xi^{(n)}) \right) \quad (t = 1, \ldots, T).$$

As $\mathcal{F}_t(\xi^{(n)}) \subseteq \mathcal{F}_t(\xi)$ holds for every $n \in \mathbb{N}$, we conclude

$$\hat{\mathcal{F}}_t \subseteq \mathcal{F}_t(\xi) \quad (\forall t = 1, \ldots, T)$$

due to the convergence of the sequence $(\xi^{(n)})$ to ξ in L_r. Furthermore, the filtration distance $D_f(\xi, \xi^{(n)})$ allows the representation

$$D_f(\xi, \xi^{(n)}) = \inf_{x \in S(\xi)} \sum_{t=2}^{T-1} \left\| x_t - \mathbb{E}[x_t | \mathcal{F}_t(\xi^{(n)})] \right\|_{r'}$$

due to (14.7). As the σ-fields $\mathcal{F}_t(\xi^{(n)})$ are nondecreasing with respect to n, we obtain for each $x \in S(\xi)$ and each $t = 2, \ldots, T$

$$\left\| x_t - \mathbb{E}[x_t | \mathcal{F}_t(\xi^{(n)})] \right\|_{r'} = \left\| \mathbb{E}[x_t | \hat{\mathcal{F}}_t] - \mathbb{E}[x_t | \mathcal{F}_t(\xi^{(n)})] \right\|_{r'} \longrightarrow 0$$

as $n \to \infty$ by classical convergence results of conditional expectations (e.g., Fetter (1977)). This completes the proof. □

The result remains true if the assumption that the σ-fields $\mathcal{F}_t(\xi^{(n)})$ are nondecreasing is replaced by the slightly weaker condition

$$\sigma\left(\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}\mathcal{F}_t(\xi^{(n)})\right) = \sigma\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\mathcal{F}_t(\xi^{(n)})\right)$$

for every $t = 2, \ldots, T$ (Fetter 1977). Next we show that ξ-adapted approximations can be obtained by certain discretization techniques.

14.3.1 Convergence of Discretizations

Let Ξ_t denote the closed subset $\operatorname{supp}(\xi_t)$ of $\mathbb{R}^d$ and $\Xi^t = \times_{\tau=1}^{t}\Xi_\tau$ for every $t = 1, \ldots, T$. Now, we consider ξ to be given on the probability space $(\Xi, \mathcal{B}(\Xi), P)$, where $\Xi = \Xi^T$ and $\mathcal{B}(\Xi)$ is the σ-field of Borel subsets of the sample space Ξ. Furthermore, the σ-fields $\mathcal{F}_t(\xi)$ are of the form

$$\mathcal{F}_t(\xi) = \left\{\xi^{-1}(B_t \times \Xi_{t+1} \times \cdots \times \Xi_T) : B_t \in \mathcal{B}(\Xi^t)\right\}. \tag{14.14}$$

14.3.1.1 Decomposition of the Sample Space

We aim at approximating the stochastic process ξ by a certain sequence of discrete processes, i.e., by processes having a finite number of scenarios. The approach is based on finite decompositions of the sample space Ξ. Let us consider a sequence

$$\mathcal{D}_t^{(n)} \subset \mathcal{B}(\Xi_t), \quad n \in \mathbb{N}.$$

It will be called a *sequence of finite segmentations* of Ξ_t if the following conditions are satisfied:

(C1) The elements of $\mathcal{D}_t^{(n)}$ are pairwise disjoint for all $n \in \mathbb{N}$,
(C2) $\mathcal{D}_t^{(n)}$ is finite and it holds $P_t\left(\cup_{D_t \in \mathcal{D}_t^{(n)}} D_t\right) = 1, n \in \mathbb{N}$.
(C3) For $\delta_{t,n} := \sup\left\{|\xi_t - \tilde{\xi}_t| : D_t \in \mathcal{D}_t^{(n)}, \xi_t, \tilde{\xi}_t \in D_t, |\xi_t|, |\tilde{\xi}_t| \le n\right\}$
it holds $\lim_{n\to\infty}\delta_{t,n} = 0$ for every $t = 1, \ldots, T$.

Conditions (C1) and (C2) ensure that the sets $\mathcal{D}_t^{(n)}$ define finite partitions of the sample space Ξ_t for every $n \in \mathbb{N}$ such that P-almost every element of Ξ_t can be associated with a unique set in $\mathcal{D}_t^{(n)}$. Condition (C3) says that the partition sets get arbitrarily small uniformly within increasing balls of radii n in Ξ_t.

Next we define a sequence of finite segmentations in Ξ by

$$\mathcal{D}^{(n)} := \left\{D_1 \times \cdots \times D_T : D_t \in \mathcal{D}_t^{(n)}, t = 1, \ldots, T\right\}, \quad n \in \mathbb{N}. \tag{14.15}$$

14.3.1.2 Discretization of the Stochastic Process

Using the sequence $\mathcal{D}^{(n)}$ we will define a sequence of approximate stochastic processes $\xi^{(n)} = (\xi_1^{(n)}, \ldots, \xi_T^{(n)})$. To this end, we select *nonanticipative* elements

$$\hat{\xi}_t^{D_1, \ldots, D_t, n} \in D_t \quad \text{with} \quad |\hat{\xi}_t^{D_1, \ldots, D_t, n}| \le C \max\{1, \inf\{|y_t| : y_t \in D_t\}\} \quad (14.16)$$

for every $n \in \mathbb{N}$, $t \in \{1, \ldots, T\}$ and every set $D_1 \times \cdots \times D_T \in \mathcal{D}^{(n)}$, where the boundedness condition in (14.16) has to be satisfied for some fixed constant $C \ge 1$. In this way we obtain a well-defined scenario

$$\hat{\xi}_{D_1 \times \cdots \times D_T}^{(n)} := \left(\hat{\xi}_1^{D_1, n}, \ldots, \hat{\xi}_T^{D_1, \ldots, D_T, n} \right)$$

for every $n \in \mathbb{N}$ and $D_1 \times \cdots \times D_T \in \mathcal{D}^{(n)}$ and define an approximate stochastic process by

$$\xi^{(n)}(\omega) := \hat{\xi}_{D_1 \times \cdots \times D_T}^{(n)} \text{ if } \omega \in \xi^{-1}(D_1 \times \cdots \times D_T) \quad (14.17)$$

and have $\xi^{(n)}$ well-defined on Ω $\mathbb{P}$-almost surely. The stochastic processes $\xi^{(n)}$ are approximations of ξ in the following sense.

Proposition 3 *Let* $\xi \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ *and (C1), (C2), and (C3) be satisfied for each* $t = 1, \ldots, T$. *Then each stochastic process* $\xi^{(n)}$ *defined by (14.17) is* ξ-*adapted and it holds*

$$\lim_{n \to \infty} \|\xi - \xi^{(n)}\|_r = 0.$$

If, in addition, (A1), (A2), and (A3) are satisfied, $r' \in [1, +\infty)$, $S(\xi)$ *is nonempty, and* $\mathcal{D}_t^{(n+1)}$ *is a refinement of* $\mathcal{D}_t^{(n)}$ *for each* $n \in \mathbb{N}$, *one has*

$$\lim_{n \to \infty} D_f(\xi, \xi^{(n)}) = 0.$$

Proof Due to the construction of $\xi^{(n)}$, the sets

$$\left\{ \xi^{-1}(D_1 \times \cdots \times D_t \times \Xi_{t+1} \times \cdots \times \Xi_T) : D_\tau \in \mathcal{D}_\tau^{(n)}, \tau = 1, \ldots, t \right\}$$

generate the σ-fields $\mathcal{F}_t(\xi^{(n)})$. Thus, it holds $\mathcal{F}_t(\xi^{(n)}) \subseteq \mathcal{F}_t(\xi)$ according to (14.14) for every $n \in \mathbb{N}$ and $t = 1, \ldots, T$.

Next we show the L_r-convergence of $(\xi^{(n)})$. To this end, let

$$B_{\gamma_n}(0) := \left\{ y \in \Xi : \max_{t=1, \ldots, T} |y_t| \le \gamma_n \right\}$$

the closed ball in Ξ around the origin with radius

$$\gamma_n := n - \max_{t=1\ldots,T} \delta_{t,n}\,.$$

Then, by using (C1) and (C2) we obtain

$$\begin{aligned}
\left\| \xi - \xi^{(n)} \right\|_r^r &= \int_\Omega \left| \xi(\omega) - \xi^{(n)}(\omega) \right|^r \mathbb{P}(d\omega) \\
&= \sum_{D_1 \times \cdots \times D_T \in \mathcal{D}^{(n)}} \int_{D_1 \times \cdots \times D_T} |\xi - \hat{\xi}^{(n)}_{D_1,\ldots,D_T}|^r \, P(d\xi) \\
&\leq c_r \sum_{D_1 \times \cdots \times D_T \in \mathcal{D}^{(n)}} \int_{D_1 \times \cdots \times D_T} \sum_{t=1}^{T} |\xi_t - \hat{\xi}^{D_1,\ldots,D_t,n}_t|^r \, P(d\xi),
\end{aligned}$$

where c_r is a suitable (norm equivalence) constant. Splitting the integration interval with respect to $B_{\gamma_n}(0)$ and its complement, using (C3) and the boundedness condition (14.16) allows to estimate

$$\begin{aligned}
\left\| \xi - \xi^{(n)} \right\|_r^r &\leq c_r \sum_{t=1}^{T} \delta_{t,n}^r + c_r \int_{\Xi \setminus B_{\gamma_n}(0)} \sum_{t=1}^{T} (|\xi_t|(1+C))^r \, P(d\xi) \\
&\leq \hat{C} \left(\max_{t=1,\ldots,T} \delta_{t,n}^r + \int_{\Xi \setminus B_{\gamma_n}(0)} |\xi|^r \, P(d\xi) \right),
\end{aligned}$$

where $\hat{C} > 0$ is some constant not depending on n. Because both summands of the last estimate tend to zero whenever n tends to infinity, the first part is proved. The second part is a consequence of Proposition 2. $\square$

Proposition 3 and Theorem 1 provide a convergence result for discretizations of (14.1), namely,

$$\lim_{n \to \infty} v(\xi^{(n)}) = v(\xi)\,. \tag{14.18}$$

Of course, this is not surprising when recalling the convergence results for discretizations of multi-stage stochastic programs in Pennanen (2005, 2009).

To determine the probabilities of the scenarios of $\xi^{(n)}$ for some $n \in \mathbb{N}$, one has to compute

$$P(D_1 \times \cdots \times D_T) \quad \text{for every} \quad D_1 \times \cdots \times D_T \in \mathcal{D}^{(n)}.$$

This might be difficult in general. However, if additional structure on ξ is available, the discretization scheme may be adapted such that the probabilities are computationally accessible. For example, let the stochastic process ξ be driven by a finite number of mutually independent $\mathbb{R}^{d_t}$-valued random variables z_t with probability distributions P_t, $t = 2, \ldots, T$, i.e.,

$$\xi_t = g_t(\xi_1, \ldots, \xi_{t-1}, z_t),$$

where the g_t, $t = 2, \ldots, T$, denote certain measurable functions from $\mathbb{R}^{(t-1)d} \times \mathbb{R}^{d_t}$ to $\mathbb{R}^d$ (see, e.g., Kuhn (2005) and Pennanen (2009)). Then there exists a measurable function G such that $\xi = G(z_2, \ldots, z_T)$. If $\mathcal{D}_t^{(n)}$ is now a partition of the support of z_t in $\mathbb{R}^{d_t}$, $t = 2, \ldots, T$, then $\xi^{(n)}$ may be defined by

$$\xi_t^{(n)} = g_t(\xi_1^{(n)}, \ldots, \xi_{t-1}^{(n)}, z_t^{(n)}),$$

where $z_t^{(n)}$, $t = 2, \ldots, T$, has a finite number of given scenarios in every $D_t \in \mathcal{D}_t^{(n)}$. The probability distribution of $\xi^{(n)}$ is then known if $P_t(D_t)$ can be computed for all $D_t \in \mathcal{D}_t^{(n)}$, $t = 2, \ldots, T$. This covers situations, where ξ is a Gaussian process or is given by certain time series models.

14.3.2 Convergence of Discrete and Empirical Measures

Our next result deals with convergence properties of discrete probability distributions.

Proposition 4 *Let P be a probability distribution on $\mathbb{R}^{Td}$ supported by a finite set of scenarios $\Xi = \{\xi^1, \ldots, \xi^N\} \subseteq \mathbb{R}^{Td}$ with positive probabilities $p_i := P(\{\xi^i\})$. Moreover, let $(P^{(n)})_{n \in \mathbb{N}}$ be a sequence of probability distributions on $\mathbb{R}^{Td}$ supported by Ξ with probabilities $p_i^{(n)} := P^{(n)}(\{\xi^i\})$ such that*

$$\lim_{n \to \infty} p_i^{(n)} = p_i, \quad i = 1, \ldots, N.$$

Then there exist random variables ξ and $(\xi^{(n)})_{n \in \mathbb{N}}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ having probability distributions P and $P^{(n)}$, $n \in \mathbb{N}$, respectively, such that

(i) $\|\xi - \xi^{(n)}\|_r \to 0$ $(n \to \infty)$ *and*

(ii) $D_f^*(\xi, \xi^{(n)}) = \sup\limits_{x \in \mathcal{B}_\infty} \sum\limits_{t=2}^T \left\| \mathbb{E}[x_t | \mathcal{F}_t(\xi)] - \mathbb{E}[x_t | \mathcal{F}_t(\xi^{(n)})] \right\|_{r'} \to 0$ $(n \to \infty)$,

for all $1 \le r < \infty$ and $1 \le r' < \infty$, where $\mathcal{B}_\infty$ is given by

$$\mathcal{B}_\infty := \{x = (x_1, \ldots, x_T) \in L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m) : |x_t(\omega)| \le 1, \ \mathbb{P}\text{-a.s.}\} \quad (14.19)$$

and denotes the set of all essentially bounded functions.

Proof The sequence $(P^{(n)})_{n \in \mathbb{N}}$ converges weakly to P on $\mathbb{R}^{Td}$. Hence, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and there exist $\mathbb{R}^{Td}$-valued random variables ξ and $\xi^{(n)}$, $n \in \mathbb{N}$, defined on it with probability distributions P and $P^{(n)}$, $n \in \mathbb{N}$, respectively, such that it holds

$$\xi^{(n)} \to \xi \quad \mathbb{P}\text{-a.s.}$$

(see, e.g., Dudley (1989), theorem 11.7.1). Since the random variables are supported by the finite set $\varXi$, almost sure convergence also implies convergence in the rth mean, i.e.,

$$\lim_{n \to \infty} \|\xi - \xi^{(n)}\|_r = 0 \,.$$

It remains to show $\lim_{n \to \infty} D_f^*(\xi, \xi^{(n)}) = 0$. To this end, we introduce partitions $\{E_{tk}\}_{k \in I_t}$ and $\{E_{tk}^{(n)}\}_{k \in I_t}$ in $\varOmega$, which generate the σ-fields $\mathcal{F}_t(\xi)$ and $\mathcal{F}_t(\xi^{(n)})$, respectively. Let

$$E_{tk} := \{\omega \in \varOmega : (\xi_1, \ldots, \xi_t)(\omega) = (\xi_1^k, \ldots, \xi_t^k)\}, \quad k \in I_t,$$
$$E_{tk}^{(n)} := \{\omega \in \varOmega : (\xi_1^{(n)}, \ldots, \xi_t^{(n)})(\omega) = (\xi_1^k, \ldots, \xi_t^k)\}, \quad k \in I_t,$$

where $I_t \subseteq \{1, \ldots, N\}$ denotes the index set of distinguishable scenarios at time $t = 2, \ldots, T$. We set

$$\bar{E}_{tk}^{(n)} := E_{tk} \cap E_{tk}^{(n)} \quad \text{and} \quad \bar{\varOmega}_t^{(n)} := \varOmega \setminus \cup_{k \in I_t} \bar{E}_{tk}^{(n)},$$

and observe that

$$\int_{\varOmega} |\xi - \xi^{(n)}|^r \mathbb{P}(d\omega) \geq \int_{E_{tk} \setminus \bar{E}_{tk}^{(n)}} |\xi - \xi^{(n)}|^r \mathbb{P}(d\omega) + \int_{E_{tk}^{(n)} \setminus \bar{E}_{tk}^{(n)}} |\xi - \xi^{(n)}|^r \mathbb{P}(d\omega)$$
$$\geq C_{\min} \left(\mathbb{P}(E_{tk} \setminus \bar{E}_{tk}^{(n)}) + \mathbb{P}(E_{tk}^{(n)} \setminus \bar{E}_{tk}^{(n)}) \right),$$

where $C_{\min} := \min \left\{ |\xi^i - \xi^j|^r : i, j \in \{1, \ldots, N\}, \xi^i \neq \xi^j \right\}$ denotes the minimal distance between two varying scenarios. Hence, by the L_r-convergence of $\xi^{(n)}$ we conclude for all $k \in I_t$ that $\mathbb{P}(E_{tk} \setminus \bar{E}_{tk}^{(n)})$ and $\mathbb{P}(E_{tk}^{(n)} \setminus \bar{E}_{tk}^{(n)})$ tend to zero whenever n tends to infinity. Moreover, the latter implies that $\mathbb{P}(E_{tk}^{(n)})$ as well as $\mathbb{P}(\bar{E}_{tk}^{(n)})$ converge to $\mathbb{P}(E_{tk})$. Let $p_{tk} = \mathbb{P}(E_{tk})$ and $p_{tk}^{(n)} = \mathbb{P}(E_{tk}^{(n)})$. Then we have for any $x \in \mathcal{B}_\infty$

$$\left\| \mathbb{E}[x_t | \mathcal{F}_t(\xi)] - \mathbb{E}[x_t | \mathcal{F}_t(\xi^{(n)})] \right\|_{r'}^{r'} = \int_{\varOmega} \left| \mathbb{E}[x_t | \mathcal{F}_t(\xi)] - \mathbb{E}[x_t | \mathcal{F}_t(\xi^{(n)})] \right|^{r'} \mathbb{P}(d\omega)$$
$$= \sum_{k \in I_t} \int_{\bar{E}_{tk}^{(n)}} \left| \mathbb{E}[x_t | \mathcal{F}_t(\xi)] - \mathbb{E}[x_t | \mathcal{F}_t(\xi^{(n)})] \right|^{r'} \mathbb{P}(d\omega)$$
$$+ \int_{\bar{\varOmega}_t^{(n)}} \left| \mathbb{E}[x_t | \mathcal{F}_t(\xi)] - \mathbb{E}[x_t | \mathcal{F}_t(\xi^{(n)})] \right|^{r'} \mathbb{P}(d\omega)$$

$$\leq \sum_{k \in I_t} \mathbb{P}(\bar{E}_{tk}^{(n)}) \left| \frac{\int_{E_{tk}} x_t \mathbb{P}(d\omega)}{p_{tk}} - \frac{\int_{E_{tk}^{(n)}} x_t \mathbb{P}(d\omega)}{p_{tk}^{(n)}} \right|^{r'} + 2\mathbb{P}(\bar{\Omega}_t^{(n)})$$

$$\leq \sum_{k \in I_t} \mathbb{P}(\bar{E}_{tk}^{(n)}) \left(\left| \frac{\int_{\bar{E}_{tk}^{(n)}} x_t (p_{tk}^{(n)} - p_{tk})\mathbb{P}(d\omega)}{p_{tk}^{(n)} p_{tk}} \right| \right.$$

$$+ \left. \left| \frac{\int_{E_{tk} \setminus \bar{E}_{tk}^{(n)}} p_{tk}^{(n)} x_t \mathbb{P}(d\omega)}{p_{tk}^{(n)} p_{tk}} \right| + \left| \frac{\int_{E_{tk}^{(n)} \setminus \bar{E}_{tk}^{(n)}} p_{tk} x_t \mathbb{P}(d\omega)}{p_{tk}^{(n)} p_{tk}} \right| \right)^{r'}$$

$$+ 2\mathbb{P}(\bar{\Omega}_t^{(n)})$$

$$\leq \sum_{k \in I_t} \mathbb{P}(\bar{E}_{tk}^{(n)}) \left(\frac{\mathbb{P}(\bar{E}_{tk}^{(n)})|p_{tk}^{(n)} - p_{tk}|}{p_{tk}^{(n)} p_{tk}} + \frac{\mathbb{P}(E_{tk} \setminus \bar{E}_{tk}^{(n)})}{p_{tk}} \right.$$

$$+ \left. \frac{\mathbb{P}(E_{tk}^{(n)} \setminus \bar{E}_{tk}^{(n)})}{p_{tk}^{(n)}} \right)^{r'} + 2\mathbb{P}(\bar{\Omega}_t^{(n)}),$$

where we used the almost sure boundedness $|x_t| \leq 1$. The latter estimate does not depend on x and due to the fact that all summands tend to zero, $D_f^*(\xi, \xi^{(n)})$ converges to 0.   □

Proposition 4 will be used in the proof of Theorem 3 to compare two stochastic processes having identical scenarios, but different probabilities.

14.3.2.1 Empirical Distributions and Sampling

Let ξ be a Ξ-valued stochastic process defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with induced distribution P. Furthermore, let $(\xi^k)_{k \in \mathbb{N}}$ be a sequence of independent and identically distributed Ξ-valued random variables on some probability space $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*)$ such that $P = \mathbb{P}^* \xi_1^{-1}$. We consider the random empirical measures

$$P^{(k)}(\omega^*)(B) := \frac{1}{k} \sum_{j=1}^{k} \delta_{\xi^j(\omega^*)}(B), \qquad n \in \mathbb{N}, \ \omega^* \in \Omega^*, \ B \in \mathcal{B}(\Xi), \quad (14.20)$$

where δ_z denotes the probability measure on Ξ placing unit mass at $z \in \Xi$. Then the sequence $(P^{(k)}(\omega^*))$ converges $\mathbb{P}^*$-almost surely to P in the sense of weak convergence (see, e.g., Dudley (1989), chapter 11.4). The portmanteau theorem (e.g., Dudley (1989), theorem 11.1.1) offers several equivalent characterizations of weak convergence, one of them is recalled in the following for later use.

Proposition 5 *Let* $(P^{(k)})$ *be a sequence of empirical measures according to (14.20). Then it holds*

$$\lim_{k \to \infty} P^{(k)}(\omega^*)(B) = P(B), \qquad \text{for all } B \in \mathcal{B}(\varXi) \text{ with } P(\partial B) = 0,$$

for $\mathbb{P}^*$-almost every $\omega^* \in \Omega^*$, where ∂B denotes the (topological) boundary of the Borel set B in the space $\mathbb{R}^s$.

Proposition 5 allows to estimate probabilities of Borel sets in $\mathbb{R}^s$ empirically, e.g., by sampling. In particular, the probability of any set belonging to a finite segmentation $\mathcal{D}^{(n)}$ of $\varXi$ (see (14.15)) can be estimated by sampling if its boundary has Lebesgue measure zero and P is absolutely continuous.

14.3.3 Application to Scenario Tree Construction

The following conceptual algorithm represents a general approach to constructing scenario trees for multi-stage stochastic programs.

Algorithm 1 *Let ξ be the original $\varXi$-valued stochastic input process of the stochastic program (14.1) defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let P be the probability distribution of ξ.*

> **Step [1]:** *Determine a sequence of finite segmentations $\mathcal{D}^{(n)}$ in $\varXi$ such that assumptions (C1), (C2), and (C3) are satisfied (cf. Sect. 14.3.1) and choose a reasonably large $n \in \mathbb{N}$.*
> **Step [2]:** *Determine the empirical measure $P^{(k)}$ based on k independent and identically P-distributed random variables.*
> **Step [3]:** *Compute the probabilities $p^{(n,k)} = P^{(k)}(D_1 \times \cdots \times D_T)$ for every $D_1 \times \cdots \times D_T \in \mathcal{D}^{(n)}$ according to formula (14.20).*
> **Step [4]:** *Choose nonanticipative scenarios whose tth components belong to D_t for any $D_1 \times \cdots \times D_T \in \mathcal{D}^{(n)}$ according to (14.16).*
> **Step [5]:** *Finally, define the stochastic scenario tree process $\xi_{\text{tr}} := \xi_{\text{tr}}^{(n,k)}$ with scenarios chosen in step 4 endowed with the empirical probabilities $p^{(n,k)}$ from step 3.*

Next we study the asymptotic behavior of the approximate scenario trees $(\xi_{\text{tr}}^{(n,k)})_{n,k \in \mathbb{N}}$ constructed by Algorithm 1. Note that the parameters n and k measure the quality of the discretization $\mathcal{D}^{(n)}$ of $\varXi$ and of the empirical probabilities, respectively.

Theorem 3 *Let (A1), (A2), and (A3) be satisfied and X_1 be nonempty and bounded. Let $1 \le r' < \infty$ and assume that for some constant $C > 0$ the estimate*

$$D_{\text{f}}(\hat{\xi}, \tilde{\xi}) \le C D_{\text{f}}^*(\hat{\xi}, \tilde{\xi}) \tag{14.21}$$

holds for all $\hat{\xi}$ and $\tilde{\xi}$ in a neighborhood of ξ in L_r. Assume that the sequence $(\xi_{\text{tr}}^{(n,k)})_{n,k \in \mathbb{N}}$ is constructed by Algorithm 1. Furthermore, assume for the sequence

$(\mathcal{D}^{(n)})_{n\in\mathbb{N}}$ of partitions of Ξ that $\mathcal{D}^{(n+1)}$ is a refinement of $\mathcal{D}^{(n)}$ and that $P_t(\partial D_t) = 0$ for all $D_t \in \mathcal{D}_t^{(n)}$ $(t = 1, \ldots, T)$. Then it holds

$$\lim_{n\to\infty}\left(\lim_{k\to\infty} v(\xi_{tr}^{(n,k)})\right) = v(\xi) \quad P^*\text{-almost surely,} \qquad (14.22)$$

where $v(\xi_{tr}^{(n,k)})$ and $v(\xi)$ denote the optimal values of the stochastic program (14.1) with input $\xi_{tr}^{(n,k)}$ and ξ, respectively, and $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*)$ is the probability space on which the random empirical measures $P^{(k)}$, $k \in \mathbb{N}$, are defined.

Proof We use the constructions and notations of Section 14.3.1 and consider, for each $n \in \mathbb{N}$, the (discrete) stochastic processes $\xi^{(n)}$ possessing the scenarios $\hat{\xi}_{D_1\times\cdots\times D_T}^{(n)}$ with probabilities $P(D_1 \times \cdots \times D_T)$ for every $D_1 \times \cdots \times D_T \in \mathcal{D}^{(n)}$. Due to Proposition 3 the processes $\xi^{(n)}$ are ξ-adapted and it holds

$$\lim_{n\to\infty} \|\xi - \xi^{(n)}\|_r = 0 \quad \text{and} \quad \lim_{n\to\infty} D_f(\xi, \xi^{(n)}) = 0.$$

Hence, we obtain from (14.5) in Theorem 1

$$|v(\xi) - v(\xi^{(n)})| \le \varepsilon_n$$

for some sequence (ε_n) tending to zero as $n \to \infty$.

Let $\hat{\Omega}^*$ be the subset of Ω^* such that $P^*(\hat{\Omega}^*) = 1$ and the sequence $(P^{(k)}(\omega^*))$ of random empirical measures converges weakly for every $\omega^* \in \hat{\Omega}^*$.

Now, let $\omega^* \in \hat{\Omega}^*$ and let $\xi_{tr}^{(n,k)} = \xi_{tr}^{(n,k)}(\omega^*)$ be the process determined by Algorithm 1. We will show that for any large $n \in \mathbb{N}$ there exists $k(n) \in \mathbb{N}$ (depending on ω^*) such that

$$|v(\xi^{(n)}) - v(\xi_{tr}^{(n,k)}(\omega^*))| \le \frac{1}{n}$$

for all $k \ge k(n)$. Then the triangle inequality would imply

$$|v(\xi) - v(\xi_{tr}^{(n,k)}(\omega^*))| \le \varepsilon_n + \frac{1}{n} \qquad (k \ge k(n))$$

and, hence, the proof was complete.

To this end, let $n \in \mathbb{N}$ be sufficiently large and fixed. We consider the processes $\xi^{(n)}$ and $\xi_{tr}^{(n,k)}$ $(k \in \mathbb{N})$ and observe that both processes possess identical scenarios which do not depend on k. In general, only the probabilities

$$p_{D_1\times\cdots\times D_T}^{(n)} = P(D_1 \times \cdots \times D_T) \text{ and } p_{D_1\times\cdots\times D_T}^{(n,k)} = P^{(k)}(\omega^*)(D_1 \times \cdots \times D_T)$$

associated with scenario $\hat{\xi}_{D_1\times\cdots\times D_T}^{(n)}$ are different.

It holds $P(\partial(D_1 \times \cdots \times D_T)) \leq \sum_{t=1}^{T} P_t(\partial D_t) = 0$ since for every boundary point x of $D_1 \times \cdots \times D_T$ there exists $t \in \{1, \ldots, T\}$ such that $x_t \in \partial D_t$. Hence, Proposition 5 implies

$$p_{D_1 \times \cdots \times D_T}^{(n,k)} \to p_{D_1 \times \cdots \times D_T}^{(n)} \qquad (k \to \infty),$$

and Proposition 4 yields the existence of a common probability space such that

$$\|\xi^{(n)} - \xi_{\mathrm{tr}}^{(n,k)}\|_r \to 0 \quad \text{and} \quad D_{\mathrm{f}}^*(\xi^{(n)}, \xi_{\mathrm{tr}}^{(n,k)}) \to 0 \quad (k \to \infty).$$

Then estimate (14.21) implies that the inequality

$$D_{\mathrm{f}}(\xi^{(n)}, \xi_{\mathrm{tr}}^{(n,k)}) \leq C D_{\mathrm{f}}^*(\xi^{(n)}, \xi_{\mathrm{tr}}^{(n,k)})$$

holds for large n and k. By making use of Theorem 1 (applied to $\xi^{(n)}$ instead of ξ), we obtain

$$|v(\xi^{(n)}) - v(\xi_{\mathrm{tr}}^{(n,k)})| \leq L(\xi^{(n)})\big(\|\xi^{(n)} - \xi_{\mathrm{tr}}^{(n,k)}\|_r + D_{\mathrm{f}}^*(\xi^{(n)}, \xi_{\mathrm{tr}}^{(n,k)})\big)$$

for some constant $L(\xi^{(n)})$ and all sufficiently large $k \in \mathbb{N}$. This implies

$$|v(\xi^{(n)}) - v(\xi_{\mathrm{tr}}^{(n,k)})| \to 0 \qquad (k \to \infty)$$

and, in particular, the existence of $k(n)$ such that

$$|v(\xi^{(n)}) - v(\xi_{\mathrm{tr}}^{(n,k)})| \leq \frac{1}{n}$$

for all $k \geq k(n)$. $\square$

We note that the limits in (14.22) cannot be interchanged. This may be interpreted such that it makes no sense to choose n very large, i.e., to choose a very fine partition of Ξ if for some reason k is not sufficiently large. In Section 14.4 we will discuss a variant of the general scenario tree construction approach provided by Algorithm 1 that is based on successive scenario reduction.

14.4 Scenario Tree Construction Based on Scenario Reduction

Next we discuss a method for generating scenario trees which is developed in Heitsch and Römisch (2009a) and motivated by Algorithm 1. It is based on a procedure of successive scenario reduction and bundling steps for increasing time stages applied to a sufficiently large scenario set. The latter is typically obtained by sampling from the underlying distribution. The stage-wise reduction may be viewed

as a simultaneous realization of steps 1, 3, and 4 of Algorithm 1. Before describing the details, we briefly recall the ideas of optimal scenario reduction.

14.4.1 Optimal Scenario Reduction

The basic idea of scenario reduction consists in determining a (nearly) best approximation in terms of a suitable probability metric of the underlying discrete probability distribution by a probability measure with smaller support. The metric is associated with the stochastic programming model in a natural way such that the model behaves stable with respect to changes of the probability distribution. Such *natural* metrics are provided in Römisch (2003) for several classes of stochastic programs.

Originally, the concept of scenario reduction was developed in Dupačová et al. (2003) and Heitsch and Römisch (2003). More recently, it has been improved for two-stage models in Heitsch and Römisch (2007) and extended to mixed-integer and chance-constrained models in Henrion et al. (2008, 2009) as well as to multi-stage models in Heitsch and Römisch (2009b). The concept does not impose special conditions on the underlying probability distribution except the existence of certain moments.

Scenario reduction aims at reducing the number of scenarios in an optimal way. If ξ is a given random vector on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with finite support, i.e., represented by the scenarios ξ^i and probabilities p_i, $i = 1, \ldots, N$, then one may be interested in finding a suitable index subset $J \subset \{1, \ldots, N\}$ and a new random vector $\hat{\xi}$ supported only by the scenarios ξ^j, $j \notin J$, such that $\hat{\xi}$ is the best approximation to ξ. Here, we consider the norm $\| \cdot \|_r$ in L_r as the natural distance function.

If J is given, the best approximation to ξ can be given explicitly. To show this, let $A_i = \xi^{-1}(\xi^i) \in \mathcal{F}, i = 1, \ldots, N$. Then

$$\|\xi - \hat{\xi}\|_r^r = \sum_{i=1}^{N} \int_{A_i} |\xi^i - \xi^{j(i)}|^r \mathbb{P}(d\omega) = \sum_{i=1}^{N} p_i |\xi^i - \xi^{j(i)}|^r$$

for some mapping $j : \{1, \ldots, N\} \rightarrow \{1, \ldots, N\} \setminus J$ and the best approximation problem reads

$$\min \left\{ \sum_{i=1}^{N} p_i |\xi^i - \xi^{j(i)}|^r \,\Big|\, j : \{1, \ldots, N\} \rightarrow \{1, \ldots, N\} \setminus J \right\}. \tag{14.23}$$

Since the lower bound

$$\sum_{i=1}^{N} p_i |\xi^i - \xi^{j(i)}|^r \geq \sum_{i \in J} p_i \min_{j \notin J} |\xi^i - \xi^j|^r$$

is always valid, the minimum in (14.23) is attained for the mapping j : $\{1, \ldots, N\} \rightarrow \{1, \ldots, N\} \setminus J$ defined by

$$j(i) \in \arg \min_{j \notin J} |\xi^i - \xi^j|, \qquad i \in J.$$

Hence, the best approximation $\hat{\xi}$ is supported by the scenarios ξ^j with probabilities q_j, $j \notin J$, where

$$\|\xi - \hat{\xi}\|_r^r = \sum_{i \in J} p_i \min_{j \notin J} |\xi^i - \xi^j|^r, \qquad (14.24)$$

$$q_j = p_j + \sum_{\substack{i \in J \\ j(i)=j}} p_i. \qquad (14.25)$$

In other words, the *redistribution rule* (14.25) consists in assigning the new probability to a preserved scenario to be equal to the sum of its former probability and of all probabilities of deleted scenarios that are closest to it.

Finding the optimal index set J, say, with prescribed cardinality, such that it solves the combinatorial optimization problem

$$\min \left\{ \sum_{i \in J} p_i \min_{j \notin J} |\xi^i - \xi^j|^r : J \subseteq \{1, \ldots, N\}, |J| = N - n \right\} \quad (1 \le n < N),$$

is much more complicated. The latter problem represents a metric k-median problem which is known to be NP-hard, hence (polynomial-time) approximation algorithms and heuristics become important. Simple heuristics may be derived from formula (14.24) for the approximation error. The results are two heuristic algorithms to compute nearly optimal index sets J with given cardinality $N - n$.

Algorithm 2 *(Forward selection)*

[*Initialization*]
Set $J := \{1, \ldots, N\}$.

[*Index Selection*]
Determine an index $l \in J$ such that

$$l \in \arg \min_{u \in J} \sum_{k \in J \setminus \{u\}} p_k \min_{j \notin J \setminus \{u\}} |\xi^k - \xi^j|^r$$

and set $J := J \setminus \{l\}$. If the cardinality of J equals $N - n$ go to the termination step. Otherwise continue with a further index selection step.

[*Termination*]
Determine scenarios ξ^j, $j \notin J$, and apply the redistribution rule (14.25) for the final index set J.

Algorithm 3 *(Backward reduction)*

[*Initialization*]
Set $J := \emptyset$.

[*Index Selection*]
Determine an index $u \notin J$ such that

$$u \in \arg\min_{l \notin J} \sum_{k \in J \cup \{l\}} p_k \min_{j \notin J \cup \{l\}} |\xi^k - \xi^j|^r$$

and set $J := J \cup \{u\}$. If the cardinality of J equals $N - n$ go to the termination step. Otherwise continue with a further index selection step.

[*Termination*]
Determine scenarios ξ^j, $j \notin J$, and apply the redistribution rule (14.25) for the final index set J.

Optimal scenario reduction allows two interesting interpretations. The first is that it may actually be considered as a problem of *optimal quantization of probability measures* (in the sense of Graf and Luschgy (2000)) when applied to a discrete probability measure (with N atoms) and using the rth order Wasserstein distance (see also Heitsch and Römisch (2009a), lemma 2.1). Second, scenario reduction leads to a canonical decomposition of the sample space $\mathbb{R}^s$. To illustrate this fact assume that $\{\xi^j : j \notin J\}$ has been computed to reduce the original scenario set $\{\xi^1, \ldots, \xi^N\}$ contained in $\mathbb{R}^s$. Then the so-called *Voronoi regions* defined by

$$V(\xi^j) := \left\{ \xi \in \mathbb{R}^s : |\xi - \xi^j| < \min_{k \notin J \cup \{j\}} |\xi - \xi^k| \right\}, \quad j \notin J$$

represent disjoint subsets of $\mathbb{R}^s$. It is known that for strictly convex norms $|\cdot|$ the union of the closures of $V(\xi^j)$ cover $\mathbb{R}^s$ and the boundaries $\partial V(\xi^j)$ have Lebesgue measure λ^s zero. The latter holds for the l_p-norms with $1 < p < \infty$ and for $p = 2$ Voronoi regions are even convex (see Graf and Luschgy (2000), section 1). Voronoi regions are a suitable choice for the sets D_t in Section 14.3.1.

Figure 14.1 shows the Voronoi decomposition of the space R^2 obtained by scenario reduction starting from $N = 1\,000$ samples from the two-dimensional standard normal distribution computed with Algorithm 2.

14.4.2 Scenario Tree Construction

The idea of the tree construction method is to apply the scenario reduction techniques to a set of scenarios successively for increasing and decreasing time stages, respectively. This leads to forward or backward variants of a tree generation method that aims at recovering the original information structure approximately. Next we present a detailed description of the forward variant, the backward approach may be found in Heitsch and Römisch (2009a).
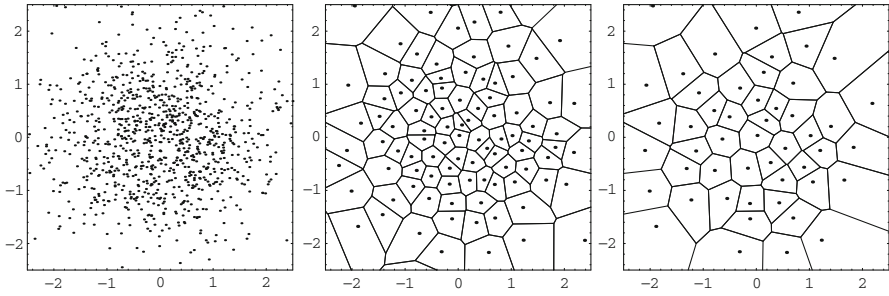
Fig. 14.1 Illustration of scenario reduction starting from 1000 sample scenarios of the two-dimensional standard normal distribution (*left*) reduced to 100 scenarios (*middle*) and further reduced to 50 scenarios (*right*). Displayed are the corresponding decompositions of $\mathbb{R}^2$ into Voronoi regions obtained by the forward selection algorithm with respect to the Euclidean norm

In the following, let $I := \{1, \ldots, N\}$ be the index set of the given set of scenarios ξ^i. Then the successive scenario reduction technique is applied to the time horizons $\{1, \ldots, t\}$ with increasing time $t \in \{1, \ldots, T\}$. It computes partitions of I of the form

$$\mathcal{C}_t := \{C_t^1, \ldots, C_t^{k_t}\}, \qquad k_t \in \mathbb{N},$$

successively such that

$$C_t^k \cap C_t^{k'} = \emptyset \quad \text{for} \quad k \neq k', \quad \text{and} \quad \bigcup_{k=1}^{k_t} C_t^k = I$$

holds for every t. The elements of a partition $\mathcal{C}_t$ are called (scenario) clusters. The following algorithm allows to generate different scenario tree processes depending on the parameter settings for the reductions in each step (Fig. 14.2).

Algorithm 4 (*Forward construction*)

[*Initialization*]
Define $C_1 = \{I\}$ and set $t := 2$.

[*Cluster computation*]
Let $C_{t-1} = \{C_{t-1}^1, \ldots, C_{t-1}^{k_{t-1}}\}$. For every $k \in \{1, \ldots, k_{t-1}\}$ apply scenario reduction to the scenario subsets $\{\xi_t^i\}_{i \in C_{t-1}^k}$ (at time t). This yields disjoint subsets of remaining and deleted scenarios I_t^k and J_t^k, respectively. Next, obtain the mappings $j_t^k : J_t^k \rightarrow I_t^k$ such that

$$j_t^k(i) \in \arg\min_{j \in I_t^k} |\xi_t^i - \xi_t^j|, \quad i \in J_t^k,$$

according to the reduction procedure (cf. Section 14.4.1). Finally, define an overall mapping $\alpha_t : I \rightarrow I$ by
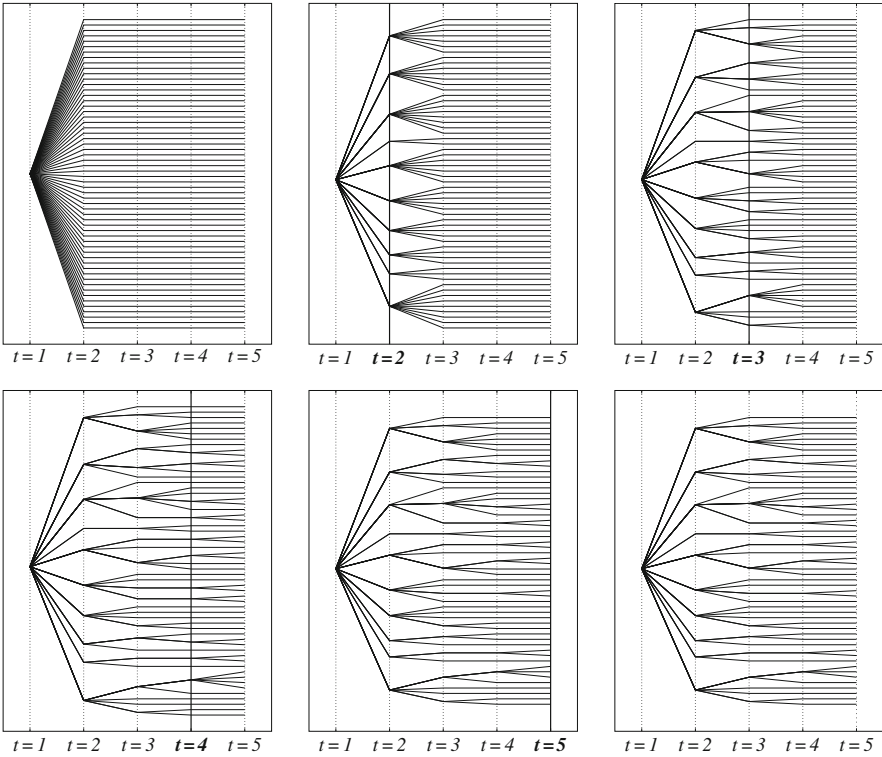
Fig. 14.2 Demonstration of the forward tree construction for an example containing $T = 5$ time periods. Displayed are the stepwise changes of the scenarios tree structure starting with a fan of individual scenarios

$$\alpha_t(i) = \begin{cases} j_t^k(i), & i \in J_t^k \text{ for some } k = 1, \ldots, k_{t-1}, \\ i & \text{otherwise.} \end{cases} \tag{14.26}$$

A new partition at t is defined now by

$$C_t := \left\{ \alpha_t^{-1}(i) \,\middle|\, i \in I_t^k, \ k = 1, \ldots, k_{t-1} \right\}$$

which is in fact a refinement of the partition C_{t-1}. If $t < T$ set $t := t + 1$ and continue with a further cluster computation step, otherwise go to the termination step.

[*Termination*]
According to the partition set C_T and mappings (14.26) define a scenario tree process ξ_{tr} supported by the scenarios

$$\xi_{\text{tr}}^k = \left(\xi_1^*, \xi_2^{\alpha_2(i)}, \ldots, \xi_t^{\alpha_t(i)}, \ldots, \xi_T^{\alpha_T(i)} \right) \quad \text{for any} \quad i \in C_T^k,$$

and probabilities $q_k := \sum\limits_{i \in C_T^k} p_i$, for each $k = 1, \ldots, k_T$.

Both heuristic algorithms from Section 14.4.1 may be used to compute the scenario reduction within every cluster computation step. According to (14.24) the error of the cluster computation step t is

$$\text{err}_t := \sum_{k=1}^{k_{t-1}} \sum_{i \in J_t^k} p_i \min_{j \in I_t^k} |\xi_t^i - \xi_t^j|^r.$$

Furthermore, as shown in Heitsch (2007, proposition 6.6), the estimate

$$\|\xi - \xi_{\text{tr}}\|_r \leq \left(\sum_{t=2}^{T} \text{err}_t \right)^{\frac{1}{r}}$$

holds for the total approximation error. The latter estimate allows to control the construction process by prescribing tolerances ε_t for err_t for every $t = 2, \ldots, T$.

14.5 Application to Electricity Management

The deregulation of energy markets has lead to an increased awareness of the need for profit maximization with simultaneous consideration of financial risk, adapted to individual risk aversion policies of market participants. Mathematical modeling of such optimization problems with uncertain input data results in large-scale stochastic programming models with a risk functional in the objective. When considering a medium-term planning horizon, one is faced with consecutive decisions based on consecutive observations, thus, the stochastic programs need to be multi-stage.

Next we report on some experiences with constructing scenario trees for a multi-stage stochastic optimization model that is tailored to the requirements of a typical German municipal power utility, which has to serve an electricity demand and a heat demand of customers in a city and its vicinity. The power utility owns a combined heat and power (CHP) facility that can serve the heat demand completely and the electricity demand partly. Further electricity can be obtained by purchasing volumes for each hour at the (day ahead) spot market of the European Energy Exchange (EEX) or by signing a supply contract for a medium-term horizon with a larger power producer. The latter possibility is suspected to be expensive, but relying on the spot market only is known to be extremely risky. Spot price risk, however, may be reduced by obtaining electricity futures at EEX. The optimization aims to maximize the mean overall revenue and, simultaneously, to minimize a risk functional on a basis of a hourly discretized optimization horizon of 1 year. Details of the optimization model can be found in Eichhorn et al. (2005).

Electricity demand and heat demand as well as spot and future prices are not known in advance, but statistical information is available due to historical observations (Fig. 14.3). A very heterogeneous statistical model is employed. It is adapted
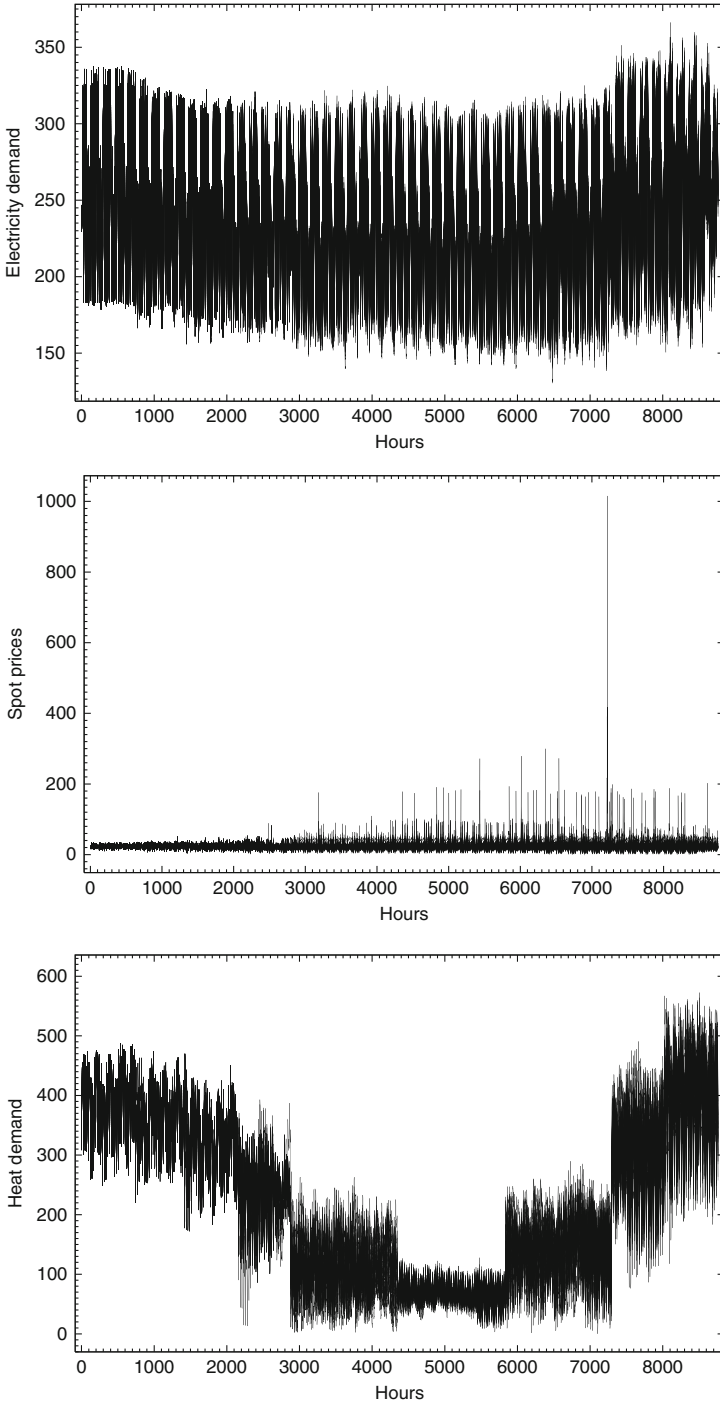
Fig. 14.3 Yearly scenario profiles of the trivariate stochastic process with components electricity demand (*top*), spot prices (*center*), and heat demand (*bottom*)

to historical data in a rather involved procedure. It consists of a cluster classification for the intra-day (demand and price) profiles and a three-dimensional time series model for the daily average values. The latter consists of deterministic trend functions and a trivariate ARMA model for the (stationary) residual time series; see Eichhorn et al. (2005) for further details. An arbitrary number of three-dimensional sample paths (scenarios) can easily be obtained by simulating white noise processes for the ARMA model and by adding on the trend functions and matched intra-day profiles from the clusters afterward. However, such a bunch of sample paths does not reflect the information structure in multi-stage stochastic optimization, i.e., it neglects the fact that information is revealed gradually over time. For this reason, finally, trivariate scenario tree processes have been computed by the approach of recursive forward scenario reduction (see Section 14.4). Table 14.1 displays the size of the three-dimensional (electricity demand, heat demand, and spot price) scenarios which serve as inputs for the tree construction (Algorithm 4). We performed a couple of test series for generating scenario trees. Due to the fact that electricity future products can only be traded monthly, branching was allowed only at the end of each month which leads to scenario trees of at most 12 stages. Because stochastic data enter both the objective and right-hand sides of the model Algorithm 4 is used with $r = r' = 2$ (cf. (14.2)). Moreover, different relative reduction levels ε_{rel} have been chosen. The relative levels are given by

$$\varepsilon_{\text{rel}} := \frac{\varepsilon}{\varepsilon_{\max}} \qquad \text{and} \qquad \varepsilon_{\text{rel},t} := \frac{\varepsilon_t}{\varepsilon_{\max}},$$

where $\varepsilon_{\max}$ is given as the best possible L_r-distance of the stochastic process represented by all scenarios and their probabilities and of one of its scenarios endowed with probability 1. The individual tolerances ε_t at branching points is computed such that

$$\varepsilon_t^r = \frac{2\varepsilon^r}{T-1}\left(q + (1-2q)\frac{t-2}{T-2}\right), \quad t = 2, \ldots, T, \qquad (14.27)$$

where $q \in [0, 1]$ is a parameter that affects the branching structure of the constructed trees. Note that a value $q < \frac{1}{2}$ generates a sequence ε_t^r with linear growth while $q > \frac{1}{2}$ results in a decreasing sequence $\varepsilon_t^r, t = 1, \ldots, T$.

Table 14.2 displays the results of our test runs with different relative reduction levels. As expected, for very small reduction levels, the reduction affects only a few scenarios. Furthermore, the number of nodes decreases considerably if the reduction level is increased. The computing times of less than 30 s already include approximately 20 s for computing distances of all scenario pairs that are needed in all calculations.

Table 14.1 Dimension of simulation scenarios

Components	Horizon	Scenarios	Time steps	Nodes
3 (trivariate)	1 year	100	8760	875,901

Table 14.2 Results of Algorithm 4 for yearly demand-price scenario trees

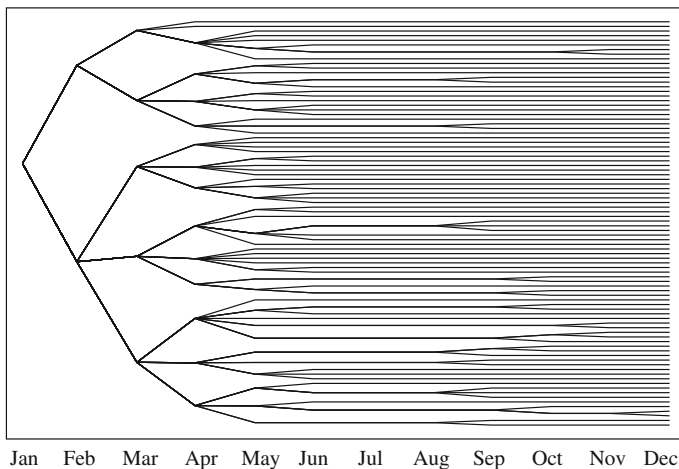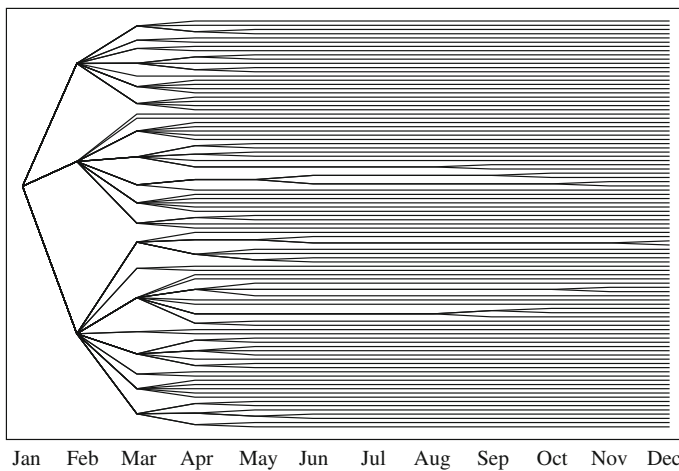| ε_{rel}	Scenarios		Nodes		Stages	Time (s)
	Initial	Tree	Initial	Tree		
0.20	100	100	875,901	775,992	4	24.53
0.25	100	100	875,901	752,136	5	24.54
0.30	100	100	875,901	719,472	7	24.55
0.35	100	97	875,901	676,416	8	24.61
0.40	100	98	875,901	645,672	10	24.64
0.45	100	96	875,901	598,704	10	24.75
0.50	100	95	875,901	565,800	9	24.74
0.55	100	88	875,901	452,184	10	24.75
0.60	100	87	875,901	337,728	11	25.89

Fig. 14.4 Yearly demand–price scenario trees obtained by Algorithm 4

Figure 14.4 illustrates the scenario trees obtained for reduction levels of 40 and 55%, respectively. Observe that in all computations the maximal number of 12 stages is not reached even at higher reduction levels. This phenomenon could be caused by the low level of heat demand during the summer period (see Fig. 14.3) such that branching occurs less frequently.

References

M. Chen and S. Mehrotra. Epi-convergent scenario generation method for stochastic problems via sparse grid. *Stochastic Programming E-Print Series*, 7, 2008.

M. A. H. Dempster. Sequential importance sampling algorithms for dynamic stochastic programming. *Zapiski Nauchnykh Seminarov POMI*, 312:94–129, 2004.

R. M. Dudley. *Real Analysis and Probability*. Chapman & Hall, New York, NY, 1989.

J. Dupačová, G. Consigli, and S. W. Wallace. Scenarios for multistage stochastic programs. *Annals of Operations Research*, 100:25–53, 2000.

J. Dupačová, N. Gröwe-Kuska, and W. Römisch. Scenario reduction in stochastic programming: An approach using probability metrics. *Mathematical Programming*, 95:493–511, 2003.

A. Eichhorn, W. Römisch, and I. Wegner. Mean-risk optimization of electricity portfolios using multiperiod polyhedral risk measures. In *IEEE St. Petersburg PowerTech Proceedings*, 2005.

H. Fetter. On the continuity of conditional expectations. *Journal of Mathematical Analysis and Applications*, 61:227–231, 1977.

K. Frauendorfer. Barycentric scenario trees in convex multistage stochastic programming. *Mathematical Programming*, 75:277–293, 1996.

S. J. Garstka and R. J-B. Wets. On decision rules in stochastic programming. *Mathematical Programming*, 7:117–143, 1974.

S. Graf and H. Luschgy. *Foundations of Quantization for Probability Distributions*, volume 1730 of *Lecture Notes in Mathematics*. Springer, Berlin, 2000.

N. Gülpinar, B. Rustem, and R. Settergren. Simulation and optimization approaches to scenario tree generation. *Journal of Economic Dynamics & Control*, 28:1291–1315, 2004.

H. Heitsch. *Stabilität und Approximation stochastischer Optimierungsprobleme*. PhD thesis, Department of Mathematics, Humboldt University, Berlin, 2007.

H. Heitsch and W. Römisch. Scenario reduction algorithms in stochastic programming. *Computational Optimization and Applications*, 24:187–206, 2003.

H. Heitsch and W. Römisch. A note on scenario reduction for two-stage stochastic programs. *Operations Research Letters*, 35:731–738, 2007.

H. Heitsch and W. Römisch. Scenario tree modeling for multistage stochastic programs. *Mathematical Programming*, 118:371–406, 2009a.

H. Heitsch and W. Römisch. Scenario tree reduction for multistage stochastic programs. *Computational Management Science*, 6:117–133, 2009b.

H. Heitsch and W. Römisch. Stability and scenario trees for multistage stochastic programs. In G. Infanger, editor, *Stochastic Programming*, 139–164. The State of the Art, In Honor of G.B. Dantzig, Springer, 2010.

H. Heitsch, W. Römisch, and C. Strugarek. Stability of multistage stochastic programs. *SIAM Journal on Optimization*, 17:511–525, 2006.

R. Henrion, C. Küchler, and W. Römisch. Discrepancy distances and scenario reduction in two-stage stochastic mixed-integer programming. *Journal of Industrial and Management Optimization*, 4:363–384, 2008.

R. Henrion, C. Küchler, and W. Römisch. Scenario reduction in stochastic programming with respect to discrepancy distances. *Computational Optimization and Applications*, 43:67–93, 2009.

R. Hochreiter and G. Ch. Pflug. Financial scenario generation for stochastic multi-stage decision processes as facility location problems. *Annals of Operations Research*, 152:257–272, 2007.

K. Høyland and S. W. Wallace. Generating scenario trees for multi-stage decision problems. *Management Science*, 47:295–307, 2001.

R. Kouwenberg. Scenario generation and stochastic programming models for asset liability management. *European Journal of Operational Research*, 134:279–292, 2001.

C. Küchler. On stability of multistage stochastic programs. *SIAM Journal on Optimization*, 19:952–968, 2008.

D. Kuhn. *Generalized Bounds for Convex Multistage Stochastic Programs, Lecture Notes in Economics and Mathematical Systems*, volume 548. Springer-Verlag Berlin Heidelberg, 2005.

C. Lemieux. *Monte Carlo and Quasi-Monte Carlo Sampling*. Springer Series in Statistics. Springer, New York, NY, 2009.

R. Mirkov and G. Ch. Pflug. Tree approximations of dynamic stochastic programs. *SIAM Journal on Optimization*, 18:1082–1105, 2007.

T. Pennanen. Epi-convergent discretizations of multistage stochastic programs. *Mathematics of Operations Research*, 30:245–256, 2005.

T. Pennanen. Epi-convergent discretizations of multistage stochastic programs via integration quadratures. *Mathematical Programming*, 116:461–479, 2009.

G. Ch. Pflug. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical Programming*, 89:251–271, 2001.

R. T. Rockafellar and R. J-B. Wets. Continuous versus measurable recourse in n-stage stochastic programming. *Journal of Mathematical Analysis and Applications*, 48:836–859, 1974.

R. T. Rockafellar and R. J-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, 1st edition, 1998. (Corr. 2nd printing 2004).

W. Römisch. Stability of stochastic programming problems. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, Volume 10 of *Handbooks in Operations Research and Management Science*. Elsevier, Amsterdam, 1st edition, chapter 8, pages 483–55, 2003.

W. Römisch. Scenario generation. In J. J. Cochran et al., editors, *Encyclopedia of Operations Research and Management Science*. Wiley Encyclopedia of Operations Research and Management Science, Wiley, Chapter 1.5.4.2, 2010.

A. Ruszczyński and A. Shapiro, editors. *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*. Elsevier, Amsterdam, 1st edition, 2003.

A. Shapiro. Inference of statistical bounds for multistage stochastic programming problems. *Mathematical Methods of Operations Research*, 58:57–68, 2003a.

A. Shapiro. Monte Carlo sampling methods. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, Volume 10 of *Handbooks in Operations Research and Management Science*. Elsevier, Amsterdam, 1st edition, chapter 6, pages 353–425, 2003b.