Stability of Stochastic Programming Problems

Werner Römisch

Institute of Mathematics, Humboldt-University Berlin, D-10099 Berlin, Germany
romisch@mathematik.hu-berlin.de

Abstract

The behaviour of stochastic programming problems is studied in case of the underlying probability distribution being perturbed and approximated, respectively. Most of the theoretical results provide continuity properties of optimal values and solution sets relative to changes of the original probability distribution, varying in some space of probability measures equipped with some convergence and metric, respectively. We start by discussing relevant notions of convergence and distances for probability measures. Then we associate a distance with a stochastic program in a natural way and derive (quantitative) continuity properties of values and solutions by appealing to general perturbation results for optimization problems. Later we show how these results relate to stability with respect to weak convergence and how certain ideal probability metrics may be associated with more specific stochastic programs. In particular, we establish stability results for two-stage and chance constrained models. Finally, we present some consequences for the asymptotics of empirical approximations and for the construction of scenario-based approximations of stochastic programs.

Key words: Stochastic programming, stability, weak convergence, probability metric, Fortet-Mourier metric, discrepancy, risk measure, two-stage, mixed-integer, chance constrained, empirical approximation, scenario reduction

1 Introduction

Stochastic programming is concerned with models for optimization problems under stochastic uncertainty that require a decision on the basis of given probabilistic information on random data. Typically, deterministic equivalents of such models represent finite-dimensional nonlinear programs whose objectives and/or constraints are given by multivariate integrals with respect to the underlying probability measure. At the modelling stage these probability measures reflect the available knowledge on the randomness at hand. This fact and the numerical challenges when evaluating the high-dimensional integrals have drawn great attention to the stability analysis of stochastic programs with respect to changes in the underlying probability measure. In this chapter we present a unified framework for such a stability analysis by regarding stochastic programs as optimization problems depending on the probability measure varying in some space of measures endowed with some distance. We give stability results both for general models and for more specific stochastic programs like two-stage and chance constrained models and include most of the proofs. Moreover, we discuss some conclusions about specific approximation procedures for stochastic programs.

To specify the stochastic programming models for our analysis, we recall that many deterministic equivalents of such models are of the form

$$\min\Big\{ \int_{\Xi} F_0(x,\xi)dP(\xi) : x \in X, \int_{\Xi} F_j(x,\xi)dP(\xi) \leq 0, j = 1, ..., d\Big\}, \qquad (1)$$

where the set $X \subseteq \mathbb{R}^m$ is closed, Ξ is a closed subset of $\mathbb{R}^s$, the functions F_j from $\mathbb{R}^m \times \Xi$ to the extended reals $\overline{\mathbb{R}}$ are random lower semicontinuous functions for $j = 0, ..., d$, and P is a Borel probability measure on Ξ.

The set X is used to describe all constraints not depending on P, and the set Ξ contains the supports of the relevant measures and provides some flexibility for formulating the models and the corresponding assumptions. We recall that F_j is a random lower semicontinuous function if its epigraphical mapping $\xi \mapsto \text{epi } F_j(\cdot,\xi) := \{(x,r) \in \mathbb{R}^m \times \mathbb{R} : F_j(x,\xi) \leq r\}$ is closed-valued and measurable, which implies, in particular, that $F_j(\cdot,\xi)$ is lower semicontinuous for each $\xi \in \Xi$ and $F_j(x,\cdot)$ is measurable for each $x \in \mathbb{R}^m$.

Although our stability analysis mainly concerns model (1) and its specifications, we also provide an approach to the stability of more general models that contain risk functionals and are of the form

$$\min\Big\{\mathbb{F}_0(P[F_0(x,\cdot)]^{-1}) : x \in X, \mathbb{F}_j(P[F_j(x,\cdot)]^{-1}) \leq 0, j = 1, ..., d\Big\}, \qquad (2)$$

where the risk functionals $\mathbb{F}_j$, $j = 0, \ldots, d$, map from suitable subsets of the set $\mathcal{P}(\mathbb{R})$ of all probability measures on $\mathbb{R}$ to $\mathbb{R}$. In general, the functionals $\mathbb{F}_j$ depend on a measure in $\mathcal{P}(\mathbb{R})$ in a more involved way than the expectation functional $\mathbb{F}_e(G) := \int_{\mathbb{R}} r dG(r)$, for which we have $\mathbb{F}_e(P[F_0(x,\cdot)]^{-1}) = \int_{\mathbb{R}} r\, dP[F_0(x,\cdot)]^{-1}(r) = \int_{\Xi} F_0(x,\xi)dP(\xi)$. Another example is the variance functional $\mathbb{F}_v(G) := \int_{\mathbb{R}} r^2 dG(r) - (\int_{\mathbb{R}} r dG(r))^2$. We also refer to the value-at-risk functional in Example 1 and to the examples in Section 2.4.

We illustrate the abstract models by the classical newsboy example (see e.g. Dupačová (1994), Example 1 in Ruszczyński and Shapiro (2003)).

Example 1 (newsboy problem)

A newsboy must place a daily order for a number x of copies of a newspaper.

He has to pay r dollars for each copy and sells a copy at c dollars, where $0 < r < c$. The daily demand ξ is random with (discrete) probability distribution $P \in \mathcal{P}(\mathbb{N})$ and the remaining copies $y(\xi) = \max\{0, x - \xi\}$ have to be removed. The newsboy might wish that the decision x maximizes his expected profit or, equivalently, minimizes his expected costs, i.e.,

$$\begin{aligned}\int_{\mathbb{R}} F_0(x,\xi)dP(\xi) &:= \int_{\mathbb{R}} [(r-c)x + c\max\{0, x-\xi\}]dP(\xi) \\ &= (r-c)x + c\sum_{k\in\mathbb{N}} \pi_k \max\{0, x-k\} \\ &= rx - cx\sum_{\substack{k\in\mathbb{N}\\ k\geq x}} \pi_k - c\sum_{\substack{k\in\mathbb{N}\\ k<x}} \pi_k k\end{aligned}$$

where π_k is the probability of demand $k \in \mathbb{N}$. The unique integer solution is the maximal $k \in \mathbb{N}$ such that $\sum_{i=k}^{\infty} \pi_i \geq \frac{r}{c}$. Another possibility is that the newsboy wishes to maximize his profit and, at the same time, to minimize his risk costs cs where s bounds the number $y(\xi)$ of copies that remain with probability p. The minimal s corresponds to his value-at-risk at level p. The resulting stochastic program reads

$$\min_{x\in\mathbb{R}_+} \Big\{(r-c)x + c \inf\{s \in \mathbb{R}_+ : P(y(\xi) \leq s) \geq p\}\Big\}.$$

The latter program is equivalent to the chance constrained model

$$\min_{(x,s)\in\mathbb{R}^2_+} \Big\{(r-c)x + cs : \sum_{\substack{k\in\mathbb{N}\\ x-s\leq k}} \pi_k \geq p\Big\} \tag{3}$$

whose unique integral solution is $(k, 0)$ with the maximal $k \in \mathbb{N}$ such that $\sum_{i=k}^{\infty} \pi_i \geq p$. Hence, the minimum risk solution is more pessimistic than the minimal expected cost solution if $\frac{r}{c} < p < 1$, i.e., if the newsboy wants to be sure with high probability that no copies of the newspaper remain.

However, the inherent difficulty of all these approaches is that the newsboy does not know the probability distribution P of the demand and has to use some approximation instead. Hence, he is interested in the stability of his decision which means that it doesn't vary too much for small perturbations of the data. For instance, his decision might be based on n independent identically distributed observations ξ_i, $i = 1, \ldots, n$, of the demand, i.e., on approximating P by the empirical measure P_n (cf. Section 4.1) and, in case of minimal expected costs, on solving the approximate problem

$$\min_{x\in\mathbb{R}_+} \Big\{(r-c)x + \frac{c}{n}\sum_{i=1}^{n} \max\{0, x-\xi_i\}\Big\}. \tag{4}$$

Of course, this approach is only justified if some optimal solution x_n of the approximate problem (4) is close to some original solution for sufficiently large n. Both variants of the newsboy problem represent specific two-stage and chance constrained stochastic programs, respectively. Their discussion will be continued in the Examples 15, 19 and 54.

Throughout we will denote the set of all Borel probability measures on Ξ by $\mathcal{P}(\Xi)$, the feasible set of (1) by $\mathcal{X}(P)$, the optimal value by $\vartheta(P)$ and the (ε-approximate) solution set of (1) by $X_\varepsilon^*(P)$ and $X^*(P)$, respectively, i.e.,

$$\mathcal{X}(P) := \Big\{x \in X : \int_\Xi F_j(x,\xi)dP(\xi) \leq 0, j = 1, ..., d\Big\}, \tag{5}$$

$$\vartheta(P) := \inf\Big\{\int_\Xi F_0(x,\xi)dP(\xi) : x \in \mathcal{X}(P)\Big\}, \tag{6}$$

$$X_\varepsilon^*(P) := \Big\{x \in \mathcal{X}(P) : \int_\Xi F_0(x,\xi)dP(\xi) \leq \vartheta(P) + \varepsilon\Big\} \ (\varepsilon \geq 0), \tag{7}$$

$$X^*(P) := X_0^*(P) = \Big\{x \in \mathcal{X}(P) : \int_\Xi F_0(x,\xi)dP(\xi) = \vartheta(P)\Big\}. \tag{8}$$

In this chapter, *stability* mostly refers to continuity properties of the optimal value function $\vartheta(\cdot)$ and the (ε-approximate) solution-set mapping $X_\varepsilon^*(\cdot)$ at P, where both $\vartheta(\cdot)$ and $X_\varepsilon^*(\cdot)$ are regarded as mappings given on a set of probability measures endowed with a suitable distance. The distance has to be selected such that it allows to estimate differences of objective and constraint function values, and, that it is optimum adapted to the model at hand. Fortunately, there exists a diversity of convergence notions and metrics in probability theory and statistics that address different goals and are based on various constructions (see, e.g., Rachev (1991), van der Vaart (1998)). We will use so-called distances with ζ-structure that are given as uniform distances of expectations of functions taken from a class $\mathcal{F}$ of measurable functions from Ξ to $\mathbb{R}$, i.e.,

$$d_{\mathcal{F}}(P,Q) = \sup_{F\in\mathcal{F}} \Big|\int_\Xi F(\xi)dP(\xi) - \int_\Xi F(\xi)dQ(\xi)\Big|. \tag{9}$$

In a first step we choose the class $\mathcal{F}$ as the set $\{F_j(x,\cdot) : x \in X \cap \mathrm{cl}\,\mathcal{U}, j = 0,\ldots,d\}$, where $\mathcal{U}$ is a properly chosen open subset of $\mathbb{R}^m$, and derive some (qualitative and quantitative) stability results in the Sections 2.2 and 2.3. Such a distance forms a kind of *minimal information (m.i.) metric* for the stability of (1). Some of the corresponding results (e.g. the Theorems 5 and 9) work under quite weak assumptions on the underlying data of (1). In particular, if possible differentiability or even continuity assumptions on the functions $x \mapsto \int_\Xi F_j(x,\xi)dP(\xi)$ are avoided for the sake of generality. The approach is inspired by general perturbation results for optimization problems in Klatte

(1987,1994), Attouch and Wets (1993) and in the monographs by Bank et al. (1982), Rockafellar and Wets (1998) and Bonnans and Shapiro (2000).

Since the m.i. metrics are often rather involved and difficult to handle, we look, on the one hand, for implications of the general qualitative result on stability with respect to the topology of weak convergence. On the other hand, we look for another metric having ζ-structure by enlarging the class $\mathcal{F}$ and, hence, bounding the m.i. metric from above. Our strategy for controlling this enlargement procedure consists in adding functions to the enlarged class that share the essential analytical properties with some of the functions $F_j(x,\cdot)$. As a result of this process we obtain *ideal* metrics that are optimum adjusted to the model (1) or to a whole class of models and that enjoy pleasant properties (e.g., a duality and convergence theory). In Section 3, we show for three types of stochastic programs how such ideal metrics come to light in a natural way by revealing the analytical properties of the relevant functions $F_j(x,\cdot)$. At the same time, we obtain quantitative stability results for all models.

For two-stage models containing integer variables and for chance constrained models, the relevant functions are discontinuous and their ideal classes contain products of (locally) Lipschitzian functions and of characteristic functions of sets describing regions of continuity (see Sections 3.2 and 3.3).

When using stability results for designing or analyzing approximation schemes or estimation procedures, further properties of the function classes $\mathcal{F}$ and of the metrics may become important. For example, we derive covering numbers of certain function classes and discuss their implications on probabilistic bounds for empirical optimal values and solution sets.

The chapter is organized as follows. First Section 2 contains some prerequisites on convergences and metric distances of probability measures. This is followed by our main qualitative stability result (Theorem 5) and its conclusions on the stability with respect to weak convergence of probability measures. We continue with the quantitative stability results for solution sets of (1) (Theorems 9 and 12) and a Lipschitz continuity result (Theorem 13) for ε-approximate solution sets of convex models. We add a discussion of how to associate ideal metrics with more specific stochastic programs. Section 2 is finished by discussing the challenges and by presenting first results of a perturbation analysis for stochastic programs containing risk functionals (2). In Section 3 we consider linear two-stage, mixed-integer two-stage and linear chance constrained stochastic programs and present various perturbation results for such models. The potential of our general perturbation analysis is explained in Section 4 for two types of approximations of the underlying probability measure P. First, we consider empirical measures as nonparametric estimators of P and derive asymptotic statistical properties of values and solutions by using empirical process theory. Secondly, we discuss the optimal construction of finitely discrete measures based on probability metrics and sketch some results and heuristic algorithms for the optimal reduction of discrete measures. We conclude the chapter with some bibliographical notes on the relevant literature.

2 General Stability Results

2.1 Convergences and Metrics of Probability Measures

Let us consider the set $\mathcal{P}(\Xi)$ of all Borel probability measures with support contained in a closed subset Ξ of $\mathbb{R}^s$. We will endow the set $\mathcal{P}(\Xi)$ or some of its subsets with different convergences and distances, which are adapted to the underlying stochastic program or to a whole class of stochastic programs. The classical convergence concept in probability theory is the *weak convergence* of measures in $\mathcal{P}(\Xi)$ (see e.g. Billingsley (1968) and Dudley (1989)). A sequence (P_n) in $\mathcal{P}(\Xi)$ is said to converge weakly to $P \in \mathcal{P}(\Xi)$, shortly $P_n \xrightarrow{w} P$, if

$$\lim_{n\to\infty} \int_{\Xi} g(\xi)dP_n(\xi) = \int_{\Xi} g(\xi)dP(\xi) \tag{10}$$

holds for each g in the space $C_{\mathrm{b}}(\Xi)$ of bounded continuous functions from Ξ to $\mathbb{R}$. It is well known that the topology τ_w of weak convergence is metrizable (e.g. by the bounded Lipschitz metric (11)) and that $P_n \xrightarrow{w} P$ holds iff the sequence of probability distribution functions of P_n converges pointwise to the distribution function F_P of P at all continuity points of F_P. Another important property of weak convergence is the continuous mapping theorem: If $P_n \xrightarrow{w} P$ and $g : \Xi \to \mathbb{R}$ is measurable, bounded and P-continuous, i.e., $P(\{\xi \in \Xi : g \text{ is not continuous at } \xi\}) = 0$, we have (10).
Most of the distances on (subsets of) $\mathcal{P}(\Xi)$ that will be considered are of the form $d_{\mathcal{F}}$ in (9), where $\mathcal{F}$ is a class of measurable functions from Ξ to $\mathbb{R}$, and are defined on the set $\mathcal{P}_{\mathcal{F}} := \{Q \in \mathcal{P}(\Xi) : \sup_{F\in\mathcal{F}} |\int_{\Xi} F(\xi)dQ(\xi)| < \infty\}$, where $d_{\mathcal{F}}$ is finite. A uniform distance of the form (9) is called a distance having *ζ-structure* (see Zolotarev (1983) and Rachev (1991)). Clearly, $d_{\mathcal{F}}$ does not change if the set $\mathcal{F}$ is replaced by its convex hull. It is nonnegative, symmetric and satisfies the triangle inequality, i.e., a pseudometric on $\mathcal{P}_{\mathcal{F}}$. $d_{\mathcal{F}}$ is a metric if the class $\mathcal{F}$ is rich enough to preserve that $d_{\mathcal{F}}(P,Q) = 0$ implies $P = Q$. Next we list some important examples of distances having ζ-structure, where the classes $\mathcal{F}$ range from (locally) Lipschitz continuous functions to piecewise constant functions with a prescribed structure of discontinuity sets.

Example 2 (metrics with ζ-structure)

(a) For $p = 0$ and $p \geq 1$ we introduce classes $\mathcal{F}_p(\Xi)$ of locally Lipschitz continuous functions that increase with p

$$\begin{aligned}\mathcal{F}_p(\Xi) &:= \{F : \Xi \mapsto \mathbb{R} : |F(\xi) - F(\tilde{\xi})| \leq c_p(\xi,\tilde{\xi})\|\xi - \tilde{\xi}\|, \forall \xi,\tilde{\xi} \in \Xi\},\\ \mathcal{F}_0(\Xi) &:= \mathcal{F}_1(\Xi) \cap \Big\{F \in C_b(\Xi) : \sup_{\xi\in\Xi} |F(\xi)| \leq 1\Big\}.\end{aligned}$$

Here, $\|\cdot\|$ denotes some norm on $\mathbb{R}^s$ and $c_p(\xi,\tilde{\xi}) := \max\{1, \|\xi\|, \|\tilde{\xi}\|\}^{p-1}$ for all $\xi, \tilde{\xi} \in \Xi$ and $p \geq 1$ describes the growth of the local Lipschitz constants. The corresponding distance with ζ-structure for $p = 0$ is the *bounded Lipschitz metric* (Section 11.3 of Dudley (1989))

$$\beta(P,Q) := \sup_{F \in \mathcal{F}_0(\Xi)} \Big| \int_\Xi F(\xi) dP(\xi) - \int_\Xi F(\xi) dQ(\xi) \Big| \tag{11}$$

and metrizes the weak convergence on $\mathcal{P}(\Xi)$. For $p = 1$ we arrive at the *Kantorovich metric*

$$\zeta_1(P,Q) := \sup_{F \in \mathcal{F}_1(\Xi)} \Big| \int_\Xi F(\xi) dP(\xi) - \int_\Xi F(\xi) dQ(\xi) \Big| \tag{12}$$

and for $p \geq 1$ at the *p-th order Fortet-Mourier metrics* (see Fortet and Mourier (1953) and Rachev (1991))

$$\zeta_p(P,Q) := \sup_{F \in \mathcal{F}_p(\Xi)} \Big| \int_\Xi F(\xi) dP(\xi) - \int_\Xi F(\xi) dQ(\xi) \Big| \tag{13}$$

on the set $\mathcal{P}_p(\Xi) := \{Q \in \mathcal{P}(\Xi) : \int_\Xi \|\xi\|^p dQ(\xi) < \infty\}$ of probability measures having finite p-th order absolute moments. It is known that a sequence (P_n) converges to P in $(\mathcal{P}_p(\Xi), \zeta_p)$ iff it converges weakly and

$$\lim_{n\to\infty} \int_\Xi \|\xi\|^p dP_n(\xi) = \int_\Xi \|\xi\|^p dP(\xi)$$

holds. Furthermore, the estimate

$$\Big| \int_\Xi \|\xi\|^p dP(\xi) - \int_\Xi \|\xi\|^p dQ(\xi) \Big| \leq p\zeta_p(P,Q)$$

is valid for each $p \geq 1$ and all $P, Q \in \mathcal{P}_p(\Xi)$ (Section 6 in Rachev (1991)). Hence, closeness with respect to ζ_p implies the closeness of q-th order absolute moments for $q \in [1,p]$.

(b) Let $\mathcal{B}$ denote a set of Borel subsets of Ξ and consider the class $\mathcal{F}_\mathcal{B} := \{\chi_B : B \in \mathcal{B}\}$ of their characteristic functions χ_B taking the value 1 if the argument belongs to B and 0 otherwise. The distance with ζ-structure generated by $\mathcal{F}_\mathcal{B}$ is defined on $\mathcal{P}(\Xi)$. It takes the form

$$\alpha_\mathcal{B}(P,Q) := d_{\mathcal{F}_\mathcal{B}}(P,Q) = \sup_{B \in \mathcal{B}} |P(B) - Q(B)|$$

and is called *$\mathcal{B}$-discrepancy*. The following instances play a special role in the context of stability in stochastic programming:
(b1) Let Ξ be convex and $\mathcal{B}_{\mathrm{c}}(\Xi)$ the set of all closed convex subsets of Ξ.
(b2) Let Ξ be polyhedral and $\mathcal{B}_{\mathrm{ph}_k}(\Xi)$ the set of all polyhedra being subsets of Ξ and having at most k faces.

(b3) Let $\Xi = \mathbb{R}^s$ and $\mathcal{B}_{\mathrm{h}}(\Xi)$ be the set of all closed half-spaces in $\mathbb{R}^s$.
(b4) Let $\Xi = \mathbb{R}^s$ and $\mathcal{B}_{\mathrm{K}}(\Xi) := \{(-\infty, \xi] : \xi \in \mathbb{R}^s\}$ be the set of all cells.
The corresponding distances are the *isotrope discrepancy* α_{c}, the *polyhedral discrepancy* α_{ph_k}, the *half-space discrepancy* α_{h} and the *Kolmogorov metric*. The latter metric coincides with the uniform distance of distribution functions on $\mathbb{R}^s$ and is denoted by d_{K}, i.e.,

$$d_{\mathrm{K}}(P,Q) = \alpha_{\mathcal{B}_{\mathrm{K}}}(P,Q) = \sup_{\xi \in \mathbb{R}^s} |P((-\infty,\xi]) - Q((-\infty,\xi])|.$$

A sequence (P_n) converges to P in $\mathcal{P}(\Xi)$ with respect to $\alpha_{\mathcal{B}}$, where $\mathcal{B}$ is a class of closed convex subsets of Ξ, iff (P_n) converges weakly to P and $P(\operatorname{bd} B) = 0$ holds for each $B \in \mathcal{B}$ (with $\operatorname{bd} B$ denoting the boundary of the set B).

The examples reveal some relations between the weak convergence of probability measures and their convergence with respect to a uniform metric $d_{\mathcal{F}}$ for some classes $\mathcal{F}$. Such relations have already been explored more systematically in the literature. A class $\mathcal{F}$ of measurable functions from Ξ to $\mathbb{R}$ is called a *P-uniformity class* if

$$\lim_{n\to\infty} d_{\mathcal{F}}(P_n, P) = 0 \tag{14}$$

holds for each sequence (P_n) that converges weakly to P. Necessary conditions for $\mathcal{F}$ to be a P-uniformity class are that $\mathcal{F}$ is uniformly bounded and that every function in $\mathcal{F}$ is P-continuous. Sufficient conditions are given in Billingsley and Topsøe (1967), Topsøe (1967,1977) and Lucchetti et al. (1994). For example, $\mathcal{F}$ is a P-uniformity class if it is uniformly bounded and it holds that $P(\{\xi \in \Xi : \mathcal{F} \text{ is not equicontinuous at } \xi\}) = 0$ (Topsøe (1967)). Unless $\mathcal{F}$ is uniformly bounded, condition (14) cannot be valid for any sequence (P_n) that converges weakly to P. In that case, a uniform integrability condition with respect to the set $\{P_n : n \in \mathbb{N}\}$ has to be additionally imposed on $\mathcal{F}$. The set $\mathcal{F}$ is called *uniformly integrable* with respect to $\{P_n : n \in \mathbb{N}\}$ if

$$\lim_{R\to\infty} \sup_{n\in\mathbb{N}} \sup_{F\in\mathcal{F}} \int_{F(\xi)>R} |F(\xi)| dP_n(\xi) = 0. \tag{15}$$

Note that condition (15) is satisfied if the moment condition

$$\sup_{n\in\mathbb{N}} \sup_{F\in\mathcal{F}} \int_{\Xi} |F(\xi)|^{1+\varepsilon} dP_n(\xi) < \infty \tag{16}$$

holds for some $\varepsilon > 0$ (Section 5 in Billingsley (1968)). Then the condition (14) is valid for any sequence (P_n) that converges weakly to P in $\mathcal{P}_{\mathcal{F}}$ and has the property that $\mathcal{F}$ is uniformly integrable with respect to $\{P_n : n \in \mathbb{N}\}$ if the set $\mathcal{F}^R := \{[F]_R(\cdot) := \max\{-R, \min\{F(\cdot), R\}\} : F \in \mathcal{F}\}$ of truncated functions of $\mathcal{F}$ is a P-uniformity class for large $R > 0$. Since the class $\mathcal{F}^R$ is uniformly bounded, it is a P-uniformity class if $P(\{\xi \in \Xi : \mathcal{F}^R$ is not equicontinuous at $\xi\}) = 0$. Sufficient conditions for classes of characteristic functions of convex sets to be P-uniformity classes are mentioned in Example 2(b).

2.2 Qualitative Stability

Together with the original stochastic programming problem (1) we consider a perturbation $Q \in \mathcal{P}(\Xi)$ of the probability distribution P and the perturbed model

$$\min\Big\{\int_\Xi F_0(x,\xi)dQ(\xi) : x \in X, \int_\Xi F_j(x,\xi)dQ(\xi) \le 0, j = 1, ..., d\Big\} \qquad (17)$$

under the general assumptions imposed in Section 1. To fix our setting, let $\|\cdot\|$ denote the Euclidean norm and $\langle\cdot,\cdot\rangle$ the corresponding inner product. By $\mathbb{B}$ we denote the Euclidean unit ball and by $d(x, D)$ the distance of $x \in \mathbb{R}^m$ to the set $D \subset \mathbb{R}^m$. For any nonempty and open subset $\mathcal{U}$ of $\mathbb{R}^m$ we consider the following sets of functions, elements and probability measures

$$\begin{aligned}
\mathcal{F}_\mathcal{U} &:= \{F_j(x,\cdot) : x \in X \cap \operatorname{cl}\mathcal{U}, j = 0, ..., d\},\\
\mathcal{X}_\mathcal{U}(Q) &:= \Big\{x \in X \cap \operatorname{cl}\mathcal{U} : \int_\Xi F_j(x,\xi)dQ(\xi) \le 0, j = 1, ..., d\Big\} \quad (Q \in \mathcal{P}_{\mathcal{F}_\mathcal{U}}(\Xi)),\\
\mathcal{P}_{\mathcal{F}_\mathcal{U}}(\Xi) &:= \Big\{Q \in \mathcal{P}(\Xi) : -\infty < \int_\Xi \inf_{x \in X \cap r\mathbb{B}} F_j(x,\xi)dQ(\xi) \quad \text{for each } r > 0 \text{ and}\\
&\qquad \sup_{x \in X \cap \operatorname{cl}\mathcal{U}} \int_\Xi F_j(x,\xi)dQ(\xi) < \infty \quad \text{for each } j = 0, ..., d\Big\},
\end{aligned}$$

and the pseudometric on $\mathcal{P}_{\mathcal{F}_\mathcal{U}} := \mathcal{P}_{\mathcal{F}_\mathcal{U}}(\Xi)$

$$d_{\mathcal{F}_\mathcal{U}}(P,Q) := \sup_{F \in \mathcal{F}_\mathcal{U}} \Big|\int_\Xi F(\xi)(P-Q)(d\xi)\Big| = \sup_{\substack{j=0,\dots,d\\ x \in X \cap \operatorname{cl}\mathcal{U}}} \Big|\int_\Xi F_j(x,\xi)(P-Q)(d\xi)\Big|.$$

Thus, $d_{\mathcal{F}_\mathcal{U}}$ is a distance of probability measures having ζ-structure. It is nonnegative, symmetric and satisfies the triangle inequality (see also Section 2.1). Our general assumptions and the Fatou Lemma imply that the objective function and the constraint set of (17) are lower semicontinuous on X and closed in $\mathbb{R}^m$, respectively, for each $Q \in \mathcal{P}_{\mathcal{F}_{\mathcal{U}}}(\Xi)$. Our first results provide further basic properties of the model (17).

Proposition 3 *Let $\mathcal{U}$ be a nonempty open subset of $\mathbb{R}^m$. Then the mapping $(x, Q) \mapsto \int_\Xi F_j(x,\xi)dQ(\xi)$ from $(X \cap \mathrm{cl}\,\mathcal{U}) \times (\mathcal{P}_{\mathcal{F}_{\mathcal{U}}}, d_{\mathcal{F}_{\mathcal{U}}})$ to $\overline{\mathbb{R}}$ is sequentially lower semicontinuous for each $j = 0, ..., d$.*

Proof: Let $j = 0, ..., d$, $x \in X \cap \mathrm{cl}\,\mathcal{U}$, $Q \in \mathcal{P}_{\mathcal{F}_{\mathcal{U}}}$, (x_n) be a sequence in $X \cap \mathrm{cl}\,\mathcal{U}$ such that $x_n \to x$, and (Q_n) be a sequence converging to Q in $(\mathcal{P}_{\mathcal{F}_{\mathcal{U}}}, d_{\mathcal{F}_{\mathcal{U}}})$. Then the lower semicontinuity of $F_j(\cdot,\xi)$ for each $\xi \in \Xi$ and the Fatou Lemma imply the estimate

$$\begin{aligned}\int_\Xi F_j(x,\xi)dQ(\xi) &\leq \liminf_{n\to\infty} \int_\Xi F_j(x_n,\xi)dQ(\xi) \\ &\leq \liminf_{n\to\infty} \{d_{\mathcal{F}_{\mathcal{U}}}(Q,Q_n) + \int_\Xi F_j(x_n,\xi)Q_n(d\xi)\} \\ &= \liminf_{n\to\infty} \int_\Xi F_j(x_n,\xi)Q_n(d\xi). \qquad \square\end{aligned}$$

Proposition 4 *Let $\mathcal{U}$ be a nonempty open subset of $\mathbb{R}^m$. Then the graph of the set-valued mapping $Q \mapsto \mathcal{X}_{\mathcal{U}}(Q)$ from $(\mathcal{P}_{\mathcal{F}_{\mathcal{U}}}, d_{\mathcal{F}_{\mathcal{U}}})$ into $\mathbb{R}^m$ is sequentially closed.*

Proof: Let (Q_n) be a sequence converging to Q in $(\mathcal{P}_{\mathcal{F}_{\mathcal{U}}}, d_{\mathcal{F}_{\mathcal{U}}})$ and (x_n) be a sequence converging to x in $\mathbb{R}^m$ and such that $x_n \in \mathcal{X}_{\mathcal{U}}(Q_n)$ for each $n \in \mathbb{N}$. Clearly, we have $x \in X \cap \mathrm{cl}\,\mathcal{U}$. For $j \in \{1, ..., d\}$ we obtain from Proposition 3 that the estimate

$$\int_\Xi F_j(x,\xi)dQ(\xi) \leq \liminf_{n\to\infty} \int_\Xi F_j(x_n,\xi)Q_n(d\xi) \leq 0\,.$$

and, thus, $x \in \mathcal{X}_{\mathcal{U}}(Q)$ holds. $\square$

To obtain perturbation results for (1), a stability property of the constraint set $\mathcal{X}(P)$ when perturbing the *probabilistic constraints* is needed. Consistently with the general definition of metric regularity for multifunctions (see, e.g., Rockafellar and Wets (1998)), we consider the set-valued mapping $y \mapsto \mathcal{X}_y(P)$ from $\mathbb{R}^d$ to $\mathbb{R}^m$, where

$$\mathcal{X}_y(P) = \Big\{x \in X : \int_\Xi F_j(x,\xi)dP(\xi) \leq y_j, j = 1, ..., d\Big\},$$

and say that its inverse $x \mapsto \mathcal{X}_x^{-1}(P) = \{y \in \mathbb{R}^d : x \in \mathcal{X}_y(P)\}$ from $\mathbb{R}^m$ to $\mathbb{R}^d$ is *metrically regular* at some pair $(\bar{x}, 0) \in \mathbb{R}^m \times \mathbb{R}^d$ with $\bar{x} \in \mathcal{X}(P) = \mathcal{X}_0(P)$

if there are constants $a \geq 0$ and $\varepsilon > 0$ such that it holds for all $x \in X$ and $y \in \mathbb{R}^d$ with $\|x - \bar{x}\| \leq \varepsilon$ and $\max\limits_{j=1,\ldots,d} |y_j| \leq \varepsilon$ that

$$d(x, \mathcal{X}_y(P)) \leq a \max_{j=1,\ldots,d} \max\Big\{0, \int\limits_{\Xi} F_j(x,\xi)dP(\xi) - y_j\Big\}.$$

To state our results we will need localized versions of optimal values and solution sets. We follow the concept proposed in Robinson (1987) and Klatte (1987), and set for any nonempty open set $\mathcal{U} \subseteq \mathbb{R}^m$ and any $Q \in \mathcal{P}_{\mathcal{F}_\mathcal{U}}$

$$\vartheta_\mathcal{U}(Q) = \inf\Big\{\int\limits_{\Xi} F_0(x,\xi)dQ(\xi) : x \in \mathcal{X}_\mathcal{U}(Q)\Big\},$$

$$X^*_\mathcal{U}(Q) = \Big\{x \in \mathcal{X}_\mathcal{U}(Q) : \int\limits_{\Xi} F_0(x,\xi)dQ(\xi) = \vartheta_\mathcal{U}(Q)\Big\}.$$

A nonempty set $\mathcal{S} \subseteq \mathbb{R}^m$ is called a *complete local minimizing (CLM) set* of (17) relative to $\mathcal{U}$ if $\mathcal{U} \subseteq \mathbb{R}^m$ is open and $\mathcal{S} = X^*_\mathcal{U}(Q) \subset \mathcal{U}$. Clearly, CLM sets are sets of local minimizers, and the set $X^*(Q)$ of global minimizers is a CLM set with $X^*(Q) = X^*_\mathcal{U}(Q)$ if $X^*(Q) \subset \mathcal{U}$.
Now, we are ready to state the main qualitative stability result.

Theorem 5 *Let $P \in \mathcal{P}_{\mathcal{F}_\mathcal{U}}$ and assume that*

(i) $X^*(P)$ is nonempty and $\mathcal{U} \subseteq \mathbb{R}^m$ is an open bounded neighbourhood of $X^*(P)$,
(ii) if $d \geq 1$, the function $x \mapsto \int_\Xi F_0(x,\xi)dP(\xi)$ is Lipschitz continuous on $X \cap \mathrm{cl}\,\mathcal{U}$,
(iii) the mapping $x \mapsto \mathcal{X}_x^{-1}(P)$ is metrically regular at each pair $(\bar{x}, 0)$ with $\bar{x} \in X^*(P)$.

Then the multifunction $X^*_\mathcal{U}$ from $(\mathcal{P}_{\mathcal{F}_\mathcal{U}}, d_{\mathcal{F}_\mathcal{U}})$ to $\mathbb{R}^m$ is upper semicontinuous at P, i.e., for any open set $\mathcal{O} \supseteq X^*_\mathcal{U}(P)$ it holds that $X^*_\mathcal{U}(Q) \subseteq \mathcal{O}$ if $d_{\mathcal{F}_\mathcal{U}}(P,Q)$ is sufficiently small. Furthermore, there are positive constants L and δ such that

$$|\vartheta(P) - \vartheta_\mathcal{U}(Q)| \leq L d_{\mathcal{F}_\mathcal{U}}(P,Q) \tag{18}$$

holds and $X^*_\mathcal{U}(Q)$ is a CLM set of (17) relative to $\mathcal{U}$ whenever $Q \in \mathcal{P}_{\mathcal{F}_\mathcal{U}}$ and $d_{\mathcal{F}_\mathcal{U}}(P,Q) < \delta$. In case $d = 0$, the estimate (18) is valid with $L = 1$ and for any $Q \in \mathcal{P}_{\mathcal{F}_\mathcal{U}}$.

Proof: We consider the (localized) parametric optimization problem

$$\min\Big\{f(x,Q) = \int\limits_{\Xi} F_0(x,\xi)dQ(\xi) : \; x \in \mathcal{X}_\mathcal{U}(Q)\Big\},$$

where the probability measure Q is regarded as a parameter varying in the pseudometric space $(\mathcal{P}_{\mathcal{F}_\mathcal{U}}, d_{\mathcal{F}_\mathcal{U}})$. Proposition 4 says that the graph of the multifunction $\mathcal{X}_\mathcal{U}$ from $\mathcal{P}_{\mathcal{F}_\mathcal{U}}$ to $\mathbb{R}^m$ is sequentially closed. Hence, $\mathcal{X}_\mathcal{U}$ is upper semicontinuous on $\mathcal{P}_{\mathcal{F}_\mathcal{U}}$, since $\mathrm{cl}\,\mathcal{U}$ is compact. Furthermore, we know by Proposition 3 that the function f from $(X \cap \mathrm{cl}\,\mathcal{U}) \times \mathcal{P}_{\mathcal{F}_\mathcal{U}}$ to $\overline{\mathbb{R}}$ is sequentially lower semicontinuous and finite. Let us first consider the case of $d = 0$. Since $f(\cdot, Q)$ is lower semicontinuous, $X^*_\mathcal{U}(Q)$ is nonempty for each $Q \in \mathcal{P}_{\mathcal{F}_\mathcal{U}}$. Let $x_* \in X^*(P)$, $Q \in \mathcal{P}_{\mathcal{F}_\mathcal{U}}$ and $\tilde{x} \in X^*_\mathcal{U}(Q)$. Then the estimate

$$
\begin{aligned}
|\vartheta(P) - \vartheta_\mathcal{U}(Q)| &\leq \max\Big\{ \int_\Xi F_0(x_*, \xi)(Q - P)(d\xi), \int_\Xi F_0(\tilde{x}, \xi)(P - Q)(d\xi) \Big\} \\
&\leq d_{\mathcal{F}_\mathcal{U}}(P, Q)
\end{aligned}
$$

holds. This implies that the multifunction $X^*_\mathcal{U}$ from $(\mathcal{P}_{\mathcal{F}_\mathcal{U}}, d_{\mathcal{F}_\mathcal{U}})$ to $\mathbb{R}^m$ is closed at P and, thus, upper semicontinuous at P.

In case $d \geq 1$, condition (ii) implies that the function f is even continuous on $(X \cap \mathrm{cl}\,\mathcal{U}) \times \mathcal{P}_{\mathcal{F}_\mathcal{U}}$. Then we use Berge's classical stability analysis (see Berge (1963) for topological parameter spaces and Theorem 4.2.1 in Bank et al. (1982) for metric parameter spaces) and conclude that $X^*_\mathcal{U}$ is upper semicontinuous at P if $\mathcal{X}_\mathcal{U}$ satisfies the following (lower semicontinuity) property at some pair $(\bar{x}, P)$ with $\bar{x} \in X^*(P)$:

$$
\mathcal{X}_\mathcal{U}(P) \cap B(\bar{x}, \bar{\varepsilon}) \subseteq \mathcal{X}_\mathcal{U}(Q) + a\, d_{\mathcal{F}_\mathcal{U}}(P, Q)\mathbb{B} \;\text{ whenever }\; d_{\mathcal{F}_\mathcal{U}}(P, Q) < \bar{\varepsilon}, \tag{19}
$$

where $a \geq 0$ is the corresponding constant in condition (iii), and $\bar{\varepsilon} > 0$ is sufficiently small. To establish property (19), let $\bar{x} \in X^*(P)$, and $a = a(\bar{x}) \geq 0$, $\varepsilon = \varepsilon(\bar{x}) > 0$ be the metric regularity constants from (iii). First we observe that the estimate $\int_\Xi F_j(x, \xi)(Q - P)(d\xi) \leq d_{\mathcal{F}_\mathcal{U}}(P, Q)$ holds for any $x \in X \cap \mathrm{cl}\,\mathcal{U}$, $j \in \{1, ..., d\}$ and $Q \in \mathcal{P}_{\mathcal{F}_\mathcal{U}}$. Next we choose $\bar{\varepsilon} = \bar{\varepsilon}(\bar{x})$ such that $0 < \bar{\varepsilon} < \varepsilon$ and $\bar{x} + (a+1)\bar{\varepsilon}\mathbb{B} \subseteq \mathcal{U}$. Hence, we have $x + a\bar{\varepsilon}\mathbb{B} \subseteq \mathcal{U}$ for any $x \in \bar{x} + \bar{\varepsilon}\mathbb{B}$. Let $Q \in \mathcal{P}_{\mathcal{F}_\mathcal{U}}$ be such that $d_{\mathcal{F}_\mathcal{U}}(P, Q) < \bar{\varepsilon}$. Putting $y_j = -d_{\mathcal{F}_\mathcal{U}}(P, Q)$, $j = 1, ..., d$, the above estimate implies that $\mathcal{X}_y(P) \cap \mathrm{cl}\,\mathcal{U} \subseteq \mathcal{X}_\mathcal{U}(Q)$. Due to the choice of $\bar{\varepsilon}$ we have $d(x, \mathcal{X}_y(P) \cap \mathrm{cl}\,\mathcal{U}) = d(x, \mathcal{X}_y(P))$ for any $x \in \mathcal{X}_\mathcal{U}(P) \cap (\bar{x} + \bar{\varepsilon}\mathbb{B})$, and, hence, the metric regularity condition (iii) yields the estimate

$$
\begin{aligned}
d(x, \mathcal{X}_\mathcal{U}(Q)) &\leq d(x, \mathcal{X}_y(P) \cap \mathrm{cl}\,\mathcal{U}) = d(x, \mathcal{X}_y(P)) \\
&\leq a \max_{j=1,\dots,d} \max\Big\{0, \int_\Xi F_j(x, \xi)dP(\xi) + d_{\mathcal{F}_\mathcal{U}}(P, Q)\Big\} \\
&\leq a\, d_{\mathcal{F}_\mathcal{U}}(P, Q),
\end{aligned}
$$

which is equivalent to the property (19). Hence, $X^*_\mathcal{U}$ is sequentially upper semicontinuous at P and there exists a constant $\hat{\delta} > 0$ such that $X^*_\mathcal{U}(Q) \subset \mathcal{U}$ for any $Q \in \mathcal{P}_{\mathcal{F}_\mathcal{U}}$ with $d_{\mathcal{F}_\mathcal{U}}(P, Q) < \hat{\delta}$. Thus $X^*_\mathcal{U}(Q)$ is a CLM set of (17) relative

to $\mathcal{U}$ for each such Q.
Moreover, for any $x \in \mathcal{X}_{\mathcal{U}}(Q) \cap (\bar{x} + \bar{\varepsilon}\,\mathbb{B})$ (iii) implies the estimate

$$\begin{aligned} d(x, \mathcal{X}_{\mathcal{U}}(P)) &= d(x, \mathcal{X}_0(P) \cap \mathrm{cl}\,\mathcal{U}) = d(x, \mathcal{X}_0(P)) \\ &\le a \max_{j=1,\dots,d} \max\Big\{0, \int_{\Xi} F_j(x,\xi)dP(\xi)\Big\} \\ &\le a \max_{j=1,\dots,d} \max\Big\{0, \int_{\Xi} F_j(x,\xi)dP(\xi) - \int_{\Xi} F_j(x,\xi)dQ(\xi)\Big\} \\ &\le a\, d_{\mathcal{F}_{\mathcal{U}}}(P,Q)\,, \end{aligned}$$

which is equivalent to the inclusion

$$\mathcal{X}_{\mathcal{U}}(Q) \cap (\bar{x} + \bar{\varepsilon}\,\mathbb{B}) \subseteq \mathcal{X}_{\mathcal{U}}(P) + a\, d_{\mathcal{F}_{\mathcal{U}}}(P,Q)\mathbb{B}.$$

Since $X^*(P)$ is compact, we employ a finite covering argument and arrive at two analogues of both inclusions, where a neighbourhood $\mathcal{N}$ of $X^*(P)$ appears instead of the balls $\bar{x} + \bar{\varepsilon}\,\mathbb{B}$ in their left-hand sides, and a uniform constant $\hat{a}$ appears instead of a in their right-hand sides. Moreover, there exists a uniform constant $\hat{\varepsilon} > 0$ such that the (new) inclusions are valid whenever $d_{\mathcal{F}_{\mathcal{U}}}(P,Q) < \hat{\varepsilon}$. Now, we choose $\delta > 0$ such that $\delta \le \min\{\hat{\delta}, \hat{\varepsilon}\}$ and $X^*_{\mathcal{U}}(Q) \subset \mathcal{N}$ whenever $d_{\mathcal{F}_{\mathcal{U}}}(P,Q) < \delta$.
Let $Q \in \mathcal{P}_{\mathcal{F}_{\mathcal{U}}}$ be such that $d_{\mathcal{F}_{\mathcal{U}}}(P,Q) < \delta$ and $\tilde{x} \in X^*_{\mathcal{U}}(Q) \subseteq \mathcal{X}_{\mathcal{U}}(Q) \cap \mathcal{N}$. Then there exists an element $\bar{x} \in \mathcal{X}_{\mathcal{U}}(P)$ satisfying $\|\tilde{x} - \bar{x}\| \le \hat{a}\, d_{\mathcal{F}_{\mathcal{U}}}(P,Q)$. We obtain

$$\begin{aligned} \vartheta(P) \le f(\bar{x}, P) &\le f(\tilde{x}, Q) + |f(\bar{x}, P) - f(\tilde{x}, Q)| \\ &\le \vartheta_{\mathcal{U}}(Q) + |f(\bar{x}, P) - f(\tilde{x}, P)| + |f(\tilde{x}, P) - f(\tilde{x}, Q)| \\ &\le \vartheta_{\mathcal{U}}(Q) + L_f\|\bar{x} - \tilde{x}\| + d_{\mathcal{F}_{\mathcal{U}}}(P,Q) \\ &\le \vartheta_{\mathcal{U}}(Q) + (L_f\hat{a} + 1)d_{\mathcal{F}_{\mathcal{U}}}(P,Q)\,, \end{aligned}$$

where $L_f \ge 0$ denotes a Lipschitz constant of $f(.,P)$ on $X \cap \mathrm{cl}\,\mathcal{U}$. For the converse estimate, let $\bar{x} \in X^*(P)$ and $Q \in \mathcal{P}_{\mathcal{F}_{\mathcal{U}}}$ be such that $d_{\mathcal{F}_{\mathcal{U}}}(P,Q) < \delta$. Then there exists $\tilde{x} \in \mathcal{X}_{\mathcal{U}}(Q)$ such that $\|\tilde{x} - \bar{x}\| \le \hat{a}\, d_{\mathcal{F}_{\mathcal{U}}}(P,Q)$. We conclude

$$\vartheta_{\mathcal{U}}(Q) \le f(\tilde{x}, Q) \le \vartheta(P) + |f(\tilde{x}, Q) - f(\bar{x}, P)|$$

and arrive analogously at the desired continuity property of $\vartheta_{\mathcal{U}}$ by putting $L = L_f\hat{a} + 1$. $\square$

The above proof partly parallels arguments in Klatte (1987). The most restrictive requirement in the above result is the metric regularity condition (iii). Example 40 in Section 3.3 provides some insight into the necessity of

condition (iii) in the context of chance constrained models. Criteria for the metric regularity of multifunctions are given e.g. in Section 9G of Rockafellar and Wets (1998) and in Mordukhovich (1994b). Here, we do not intend to provide a specific sufficient condition for (iii), but recall that the constraint functions $\int_\Xi F_j(\cdot,\xi)dP(\xi)$ $(j = 1,\ldots,d)$ are often nondifferentiable or even discontinuous in stochastic programming. In Section 3.3 we show how metric regularity is verified in case of chance constrained programs.
Although Theorem 5 also asserts a quantitative continuity property for optimal values, its essence consists in a continuity result for optimal values and solution sets. As a first conclusion we derive consequences for the stability of (1) with respect to the weak convergence of probability measures (cf. Section 2.1). To state our main stability result for (1) with respect to the topology of weak convergence, we need the classes $\mathcal{F}_\mathcal{U}^R$ of truncated functions of $\mathcal{F}_\mathcal{U}$ for $R > 0$ and the uniform integrability property of $\mathcal{F}_\mathcal{U}$ (see Section 2.1).

Theorem 6 *Let the assumptions of Theorem 5 for (1) be satisfied. Furthermore, let $\mathcal{F}_\mathcal{U}^R$ be a P-uniformity class for large $R > 0$ and (P_n) be a sequence in $\mathcal{P}_{\mathcal{F}_\mathcal{U}}$ that is weakly convergent to P.*
Then the sequence $(\vartheta_\mathcal{U}(P_n))$ converges to $\vartheta(P)$, the sets $X^*_\mathcal{U}(P_n)$ are CLM sets relative to $\mathcal{U}$ for sufficiently large $n \in \mathbb{N}$ and

$$\lim_{n\to\infty} \sup_{x\in X^*_\mathcal{U}(P_n)} d(x, X^*(P))) = 0$$

holds if $\mathcal{F}_\mathcal{U}$ is uniformly integrable with respect to $\{P_n : n \in \mathbb{N}\}$.

Proof: Let (P_n) be a sequence in $\mathcal{P}_{\mathcal{F}_\mathcal{U}}$ that converges weakly to P and has the property that $\mathcal{F}_\mathcal{U}$ is uniformly integrable with respect to $\{P_n : n \in \mathbb{N}\}$. Then the assumption implies (see Section 2.1)

$$\lim_{n\to\infty} d_{\mathcal{F}_\mathcal{U}}(P_n, P) = 0$$

and, hence, the result is an immediate consequence of Theorem 5. □

Compared to Theorem 5, the stability of (1) with respect to weakly convergent perturbations of P requires additional conditions on $\mathcal{F}_\mathcal{U}$. The previous theorem provides the sufficient conditions that its truncated class $\mathcal{F}_\mathcal{U}^R$ has the P-uniformity property for large $R > 0$ and that $\mathcal{F}_\mathcal{U}$ is uniformly integrable with respect to the set of perturbations. The first condition is satisfied if $\mathcal{F}_\mathcal{U}^R$ is P-almost surely equicontinuous on Ξ (cf. Section 2.1). It implies, in particular, the P-continuity of $F_j(x,\cdot)$ for each $j = 0,\ldots,d$ and $x \in X \cap \mathrm{cl}\,\mathcal{U}$. The uniform integrability condition

$$\lim_{R\to\infty} \sup_{n\in\mathbb{N}} \max_{j=0,\ldots,d} \sup_{x\in X\cap \mathrm{cl}\,\mathcal{U}} \int_{|F_j(x,\xi)|>R} |F_j(x,\xi)|dP_n(\xi) = 0 \tag{20}$$

is satisfied if the moment condition

$$\sup_{n\in\mathbb{N}} \max_{j=0,\dots,d} \sup_{x\in X\cap \mathrm{cl}\,\mathcal{U}} \int_{\Xi} |F_j(x,\xi)|^{1+\varepsilon} dP_n(\xi) < \infty \tag{21}$$

holds for some $\varepsilon > 0$. Assume, for example, that the functions F_j satisfy an estimate of the form

$$|F_j(x,\xi)| \leq C\|\xi\|^k, \quad \forall (x,\xi) \in (X \cap \mathrm{cl}\,\mathcal{U}) \times \Xi,$$

for some positive constants C, k and all $j = 0, \dots, d$ (see e.g. Sections 3.1 and 3.2). In this case, the uniform integrability condition (20) is satisfied if

$$\lim_{R\to\infty} \sup_{n\in\mathbb{N}} \int_{\|\xi\|>R} \|\xi\|^k dP_n(\xi) = 0.$$

The corresponding sufficient moment condition reads

$$\sup_{n\in\mathbb{N}} \int_{\Xi} \|\xi\|^{k+\varepsilon} dP_n(\xi) < \infty$$

for some $\varepsilon > 0$. The latter condition is often imposed in stability studies with respect to weak convergence.
The P-continuity property of each function $F_j(x,\cdot)$ and condition (20) are not needed in Theorem 5. However, the following examples show that both conditions are indispensable for stability with respect to weak convergence.

Example 7 Let $m = s = 1$, $d = 0$, $\Xi = \mathbb{R}$, $X = \mathbb{R}_-$, $F_0(x,\xi) = -\chi_{(-\infty,x]}(\xi)$ for $(x,\xi) \in \mathbb{R} \times \Xi$ and $P = \delta_0$, where δ_ξ denotes the measure that places unit mass at ξ. Then $\vartheta(P) = 1$ and $X^*(P) = \{0\}$. The sequence $(\delta_{\frac{1}{n}})$ converges weakly to P in $\mathcal{P}(\Xi)$, but it holds that $\vartheta(P_n) = 0$ for each $n \in \mathbb{N}$. This is due to the fact that, for some neighbourhood $\mathcal{U}$ of 0, the set $\{\chi_{(-\infty,x]}(\cdot) : x \in X \cap \mathrm{cl}\,\mathcal{U}\}$ is not a P-uniformity class since $P(\mathrm{bd}\,(-\infty, 0]) = P(\{0\}) = 1$.

Example 8 Let $m = s = 1$, $d = 0$, $\Xi = \mathbb{R}_+$, $X = [-1,1]$, $F_0(x,\xi) = \max\{\xi - x, 0\}$ for $(x,\xi) \in \mathbb{R} \times \Xi$ and $P = \delta_0$. Then $\vartheta(P) = 0$ and $X^*(P) = [0,1]$. Consider the sequence $P_n = (1 - \frac{1}{n})\delta_0 + \frac{1}{n}\delta_n$, $n \in \mathbb{N}$, which converges weakly to P. It holds that $\vartheta(P_n) = 1 - \frac{1}{n}$ and $X^*(P_n) = \{1\}$ for each $n \in \mathbb{N}$ and, thus, $(\vartheta(P_n))$ does not converge to $\vartheta(P)$. Here, the reason is that the class $\{\max\{\cdot - x, 0\} : x \in [-1,1]\}$ is not uniformly integrable with respect to $\{P_n : n \in \mathbb{N}\}$.

Indeed, the weak convergence of measures is a very weak condition on sequences and, hence, requires strong conditions on (1) to be stable. Many approximations of P (e.g., in Section 4.1), however, have much stronger properties than weak convergence and, hence, work under weaker assumptions than Theorem 6. To give an example, we recall that the P-continuity property of each function $F_j(x,\cdot)$ is an indispensable assumption in case of stability with respect to weak convergence, but this property is not needed when working with $d_{\mathcal{F}_\mathcal{U}}$ and with specifically adjusted ideal metrics (and the corresponding convergences of measures) in case of (mixed-integer) two-stage and chance constrained models (see Sections 3.1, 3.2 and 3.3). Consequently, we prefer to work with these distances, having in mind their relations to the topology of weak convergence.

2.3 Quantitative Stability

The main result in the previous section claims that the multifunction $X^*_\mathcal{U}(\cdot)$ is nonempty near P and upper semicontinuous at P. In order to quantify the upper semicontinuity property, a growth condition on the objective function in a neighbourhood of the solution set to the original problem (1) is needed. Instead of imposing a specific growth condition (as e.g. quadratic growth), we consider the *growth function* ψ_P defined on $\mathbb{R}_+$ by

$$\psi_P(\tau) := \min\Big\{ \int_\Xi F_0(x,\xi)dP(\xi) - \vartheta(P) : d(x, X^*(P)) \geq \tau, x \in \mathcal{X}_\mathcal{U}(P)\Big\} \quad (22)$$

of problem (1) on $\mathrm{cl}\,\mathcal{U}$, i.e., near its solution set $X^*(P)$, and the associated function

$$\Psi_P(\eta) := \eta + \psi_P^{-1}(2\eta) \quad (\eta \in \mathbb{R}_+), \quad (23)$$

where we set $\psi_P^{-1}(t) := \sup\{\tau \in \mathbb{R}_+ : \psi_P(\tau) \leq t\}$. Both functions, ψ_P and Ψ_P, depend on the data of (1) and, in particular, on P. They are lower semicontinuous on $\mathbb{R}_+$; ψ_P is nondecreasing, Ψ_P increasing and both vanish at 0 (cf. Theorem 7.64 in Rockafellar and Wets (1998)). The second main stability result establishes a quantitative upper semicontinuity property of (localized) solution sets and identifies the function Ψ_P as modulus of semicontinuity. In the convex case, it also provides continuity moduli of countable dense families of selections to solution sets.

Theorem 9 *Let the assumptions of Theorem 5 be satisfied and $P \in \mathcal{P}_{\mathcal{F}_\mathcal{U}}$. Then there exists a constant $\hat{L} \geq 1$ such that*

$$\emptyset \neq X^*_\mathcal{U}(Q) \subseteq X^*(P) + \Psi_P(\hat{L} d_{\mathcal{F}_\mathcal{U}}(P,Q))\mathbb{B} \quad (24)$$

holds for any $Q \in \mathcal{P}_{\mathcal{F_U}}$ with $d_{\mathcal{F_U}}(P,Q) < \delta$. Here, δ is the constant in Theorem 5 and Ψ_P is given by (23). In case $d = 0$, the estimate (24) is valid with $\hat{L} = 1$ and for any $Q \in \mathcal{P}_{\mathcal{F_U}}$.

Proof: Let $L > 0$, $\delta > 0$ be the constants in Theorem 5, $Q \in \mathcal{P}_{\mathcal{F_U}}$ with $d_{\mathcal{F_U}}(P,Q) < \delta$ and $\tilde{x} \in X^*_{\mathcal{U}}(Q)$. As argued in the proof of Theorem 5, there exists an element $\bar{x} \in \mathcal{X}_{\mathcal{U}}(P)$ such that $\|\tilde{x} - \bar{x}\| \leq \hat{a}\overline{\delta}$, where $\overline{\delta} := d_{\mathcal{F_U}}(P,Q)$. Let $L_P \geq 0$ denote a Lipschitz constant of the function $x \mapsto \int_\Xi F_0(x,\xi)dP(\xi)$ on $X \cap \mathrm{cl}\,\mathcal{U}$. Then the definition of ψ and Theorem 5 imply that

$$
\begin{aligned}
\overline{\delta}(1 + L_P\hat{a} + L) &\geq \overline{\delta}(1 + L_P\hat{a}) + \vartheta_{\mathcal{U}}(Q) - \vartheta(P) \\
&= \overline{\delta}(1 + L_P\hat{a}) + \int_\Xi F_0(\tilde{x},\xi)dQ(\xi) - \vartheta(P) \\
&\geq \overline{\delta}L_P\hat{a} + \int_\Xi F_0(\tilde{x},\xi)dP(\xi) - \vartheta(P) \\
&\geq \int_\Xi F_0(\bar{x},\xi)dP(\xi) - \vartheta(P) \geq \psi_P(d(\bar{x}, X^*(P))) \\
&\geq \inf_{y \in \tilde{x} + \hat{a}\overline{\delta}\mathbb{B}} \psi_P(d(y, X^*(P))) = \psi_P(d(\tilde{x}, X^*(P) + \hat{a}\overline{\delta}\mathbb{B}))\,.
\end{aligned}
$$

Hence, we obtain

$$
\begin{aligned}
d(\tilde{x}, X^*(P)) &\leq \hat{a}\overline{\delta} + d(\tilde{x}, X^*(P) + \hat{a}\overline{\delta}\mathbb{B}) \\
&\leq \hat{a}\overline{\delta} + \psi_P^{-1}(\overline{\delta}(1 + L_P\hat{a} + L)) \leq \hat{L}\overline{\delta} + \psi_P^{-1}(2\hat{L}\overline{\delta}) = \Psi_P(\hat{L}\overline{\delta}),
\end{aligned}
$$

where $\hat{L} := \max\{\hat{a}, \frac{1}{2}(1 + L_P\hat{a} + L)\} \geq 1$. In case $d = 0$, we may choose $\hat{x} = \tilde{x}$, $\hat{a} = 1$, $L = 1$, $L_P = 0$ and an arbitrary δ. This completes the proof. □

Parts of the proof are similar to arguments of Theorem 7.64 in Rockafellar and Wets (1998). Next, we briefly comment on some aspects of the general stability theorems, namely, specific growth conditions and localization issues.

Remark 10 Problem (1) is said to have *k-th order growth* at the solution set for some $k \geq 1$ if $\psi_P(\tau) \geq \gamma\tau^k$ for each small $\tau \in \mathbb{R}_+$ and some $\gamma > 0$, i.e., if

$$
\int_\Xi F_0(x,\xi)dP(\xi) \geq \vartheta(P) + \gamma d(x, X^*(P))^k
$$

holds for each feasible x close to $X^*(P)$. Then $\Psi_P(\eta) \leq \eta + (\frac{2\eta}{\gamma})^{\frac{1}{k}} \leq C\eta^{\frac{1}{k}}$ for some constant $C > 0$ and sufficiently small $\eta \in \mathbb{R}_+$. In this case, Theorem 9 provides the Hölder continuity of $X^*_{\mathcal{U}}$ at P with rate $\frac{1}{k}$. Important special cases are the linear and quadratic growth for $k = 1$ and $k = 2$, respectively.

Remark 11 In the Theorems 5 and 9 the localized optimal values $\vartheta_{\mathcal{U}}(Q)$ and solution sets $X^*_{\mathcal{U}}(Q)$ of the (perturbed) model (17) may be replaced by their global versions $\vartheta(Q)$ and $X^*(Q)$ if there exists a constant $\delta_0 > 0$ such that for each $Q \in \mathcal{P}_{\mathcal{F}_{\mathcal{U}}}$ with $d_{\mathcal{F}_{\mathcal{U}}}(P, Q) < \delta_0$ either of the following conditions is satisfied: (a) The model (17) is convex and $X^*_{\mathcal{U}}(Q)$ is a CLM set, (b) the constraint set of (17) is contained in some bounded set $\mathcal{V} \subset \mathbb{R}^m$ not depending on Q, and it holds that $\mathcal{V} \subseteq \mathcal{U}$.

In case of a fixed constraint set, i.e., $d = 0$, we derive an extension of Theorem 9 by using a probability distance that is based on divided differences of the functions $x \mapsto \int_\Xi F_0(x, \xi) d(P - Q)(\xi)$ around the solution set of (1). For some nonempty, bounded, open subset $\mathcal{U}$ of $\mathbb{R}^m$ we consider the following set of probability measures

$$\hat{\mathcal{P}}_{\mathcal{F}_{\mathcal{U}}} := \Big\{ Q \in \mathcal{P}_{\mathcal{F}_{\mathcal{U}}} : \exists C_Q > 0 \text{ such that } \int_\Xi \frac{F_0(x,\xi) - F_0(\bar{x},\xi)}{\|x - \bar{x}\|} dQ(\xi) \leq C_Q, \\ \forall x, \bar{x} \in X \cap \mathrm{cl}\,\mathcal{U}, x \neq \bar{x} \Big\}$$

and the distance

$$\hat{d}_{\mathcal{F}_{\mathcal{U}}}(P, Q) := \sup \Big\{ \int_\Xi \frac{F_0(x,\xi) - F_0(\bar{x},\xi)}{\|x - \bar{x}\|} d(P - Q)(\xi) : x, \bar{x} \in X \cap \mathrm{cl}\,\mathcal{U}, x \neq \bar{x} \Big\}$$

which is well defined and finite on $\hat{\mathcal{P}}_{\mathcal{F}_{\mathcal{U}}}$. The following result has been inspired by Section 4.4.1 in Bonnans and Shapiro (2000).

Theorem 12 *Let $d = 0$, $P \in \hat{\mathcal{P}}_{\mathcal{F}_{\mathcal{U}}}$, $X^*(P)$ be nonempty and $\mathcal{U} \subset \mathbb{R}^m$ be a bounded and open neighbourhood of $X^*(P)$. Then the estimate*

$$\sup_{x \in X^*_{\mathcal{U}}(Q)} d(x, X^*(P)) \leq (\psi^r_P)^{-1}(\hat{d}_{\mathcal{F}_{\mathcal{U}}}(P, Q))$$

is valid for any $Q \in \hat{\mathcal{P}}_{\mathcal{F}_{\mathcal{U}}}$, where $\psi^r_P(0) = 0$, $\psi^r_P(\tau) := \frac{\psi_P(\tau)}{\tau}$ for each $\tau > 0$ and $\psi_P(\cdot)$ is the growth function given by (22).
If, moreover, $(\psi^r_P)^{-1}$ is continuous at $\tau = 0$, there exists a constant $\delta > 0$ such that $X^*_{\mathcal{U}}(Q)$ is a CLM set relative to $\mathcal{U}$ whenever $\hat{d}_{\mathcal{F}_{\mathcal{U}}}(P, Q) < \delta$.
If, in particular, the original problem (1) has quadratic growth, i.e., $\psi_P(\tau) \geq \gamma\tau^2$ for some $\gamma > 0$, there exists a constant $\delta > 0$ such that the inclusion

$$\emptyset \neq X^*_{\mathcal{U}}(Q) \subseteq X^*(P) + \frac{1}{\gamma} \hat{d}_{\mathcal{F}_{\mathcal{U}}}(P, Q) \mathbb{B}$$

holds whenever $\hat{d}_{\mathcal{F}_{\mathcal{U}}}(P, Q) < \delta$.

Proof: Let $Q \in \hat{\mathcal{P}}_{\mathcal{F}_\mathcal{U}}$, $x \in X^*_\mathcal{U}(Q)$ and $\bar{x} \in X^*(P)$ be such that $\|x - \bar{x}\| = d(x, X^*(P)) > 0$. We denote $f_Q(y) := \int_\Xi F_0(y,\xi)dQ(\xi)$ for each $y \in X$, and have $f_Q(x) \leq f_Q(\bar{x})$ and $f_P(x) - f_P(\bar{x}) \geq \psi_P(d(x, X^*(P))) = \psi_P(\|x - \bar{x}\|)$. This leads to the following estimate

$$
\begin{aligned}
\psi^r_P(\|x-\bar{x}\|) &= \frac{1}{\|x-\bar{x}\|}\psi_P(\|x-\bar{x}\|) \leq \frac{1}{\|x-\bar{x}\|}(f_P(x) - f_P(\bar{x})) \\
&\leq \frac{1}{\|x-\bar{x}\|}(f_P(x) - f_Q(x) + f_Q(\bar{x}) - f_P(\bar{x})) \\
&= \frac{1}{\|x-\bar{x}\|}((f_P - f_Q)(x) - (f_P - f_Q)(\bar{x})) \\
&\leq \hat{d}_{\mathcal{F}_\mathcal{U}}(P,Q),
\end{aligned}
$$

which completes the first part. Since $\mathcal{U}$ is open, there exists an $\varepsilon > 0$ such that the ε-enlargement $\{x \in \mathbb{R}^m : d(x, X^*(P)) \leq \varepsilon\}$ of $X^*(P)$ is contained in $\mathcal{U}$. Let $\delta > 0$ be chosen such that $(\psi^r_P)^{-1}(\delta) \leq \varepsilon$. Then $d(x, X^*(P)) \leq \varepsilon$ and, thus, $x \in \mathcal{U}$ holds for each $x \in X^*_\mathcal{U}(Q)$, completing the second part.
Finally, it remains to remark that quadratic growth implies $\psi^r_P(\tau) \geq \gamma\tau$ for any $\tau > 0$ and some $\gamma > 0$. □

Compared to the estimate in Theorem 9 based on function values of the function F_0, the above bound uses divided difference information of F_0 relative to x and leads to Lipschitz-type results in case of quadratic growth.
While the growth behaviour of the objective function is important for the quantitative stability of solution sets even for convex models, the situation is much more advantageous for ε-approximate solution sets. For convex models (1) with a fixed constraint set (i.e., $d = 0$), we will see that the latter sets behave Lipschitz continuously with respect to changes of probability distributions measured in terms of the distance $d_{\mathcal{F}_\mathcal{U}}$, but for a larger set $\mathcal{U}$ compared with stability results for solution sets. To state the result, let

$$
\mathbb{D}_\rho(C,D) := \inf\{\eta \geq 0 : C \cap \rho\mathbb{B} \subset D + \eta\mathbb{B}, D \cap \rho\mathbb{B} \subset C + \eta\mathbb{B}\} \quad (25)
$$

$$
\mathbb{D}_\infty(C,D) := \inf\{\eta \geq 0 : C \subset D + \eta\mathbb{B}, D \subset C + \eta\mathbb{B}\} \quad (26)
$$

denote the ρ-distance ($\rho \geq 0$) and the Pompeiu-Hausdorff distance, respectively, of nonempty closed subsets C, D of $\mathbb{R}^m$.

Theorem 13 *Let $d = 0$, F_0 be a random lower semicontinuous convex function, X be closed convex, $P \in \mathcal{P}_{\mathcal{F}_\mathcal{U}}$ and $X^*(P)$ be nonempty and bounded. Then there exist constants $\rho > 0$ and $\bar{\varepsilon} > 0$ such that the estimate*

$$
\mathbb{D}_\infty(X^*_\varepsilon(P), X^*_\varepsilon(Q)) \leq \frac{2\rho}{\varepsilon} d_{\mathcal{F}_\mathcal{U}}(P,Q)
$$

holds for $\mathcal{U} := (\rho + \bar{\varepsilon})\mathbb{B}$ *and any* $\varepsilon \in (0, \bar{\varepsilon})$, $Q \in \mathcal{P}_{\mathcal{F}_\mathcal{U}}$ *such that* $d_{\mathcal{F}_\mathcal{U}}(P, Q) < \varepsilon$.

Proof: First we choose $\rho_0 > 0$ such that $X^*(P)$ is contained in the open ball $\mathcal{U}_{\rho_0}$ around the origin in $\mathbb{R}^m$ with radius ρ_0 and that $\vartheta(P) \geq -\rho_0 + 1$. Applying Theorem 5 with $\mathcal{U}_{\rho_0}$ as the bounded open neighbourhood of $X^*(P)$, we obtain some constant $\varepsilon_0 > 0$ such that $X^*(Q)$ is nonempty and contained in $\mathcal{U}_{\rho_0}$ and $\vartheta(Q) \geq \rho_0$ holds whenever $Q \in \mathcal{P}_{\mathcal{F}_{\mathcal{U}_{\rho_0}}}$ and $d_{\mathcal{F}_{\mathcal{U}_{\rho_0}}}(P, Q) < \varepsilon_0$. Now, let $\rho > \rho_0$, $\bar{\varepsilon} := \min\{\varepsilon_0, \rho - \rho_0, 1\}$ and $\mathcal{U} := (\rho + \bar{\varepsilon})\mathbb{B}$.
For any $Q \in \mathcal{P}_{\mathcal{F}_\mathcal{U}}$ we set again $f_Q(x) := \int_\Xi F_0(x, \xi) dQ(\xi)$ for each $x \in \mathbb{R}^m$. Furthermore, we denote by $\widehat{d}_\rho^+$ the auxiliary epi-distance of f_P and f_Q introduced in Proposition 7.61 in Rockafellar and Wets (1998):

$$\widehat{d}_\rho^+(f_P, f_Q) := \inf\{\eta \geq 0 : \inf_{y \in x + \eta\mathbb{B}} f_Q(y) \leq \max\{f_P(x), -\rho\} + \eta, \\ \inf_{y \in x + \eta\mathbb{B}} f_P(y) \leq \max\{f_Q(x), -\rho\} + \eta, \forall x \in \rho\mathbb{B}\}.$$

From Theorem 7.69 in Rockafellar and Wets (1998) we conclude that the estimate

$$\mathbb{D}_\rho(X_\varepsilon^*(P), X_\varepsilon^*(Q)) \leq \frac{2\rho}{\varepsilon} \widehat{d}_{\rho+\varepsilon}^+(f_P, f_Q)$$

is valid for $\varepsilon \in (0, \bar{\varepsilon})$ if $\widehat{d}_{\rho+\varepsilon}^+(f_P, f_Q) < \varepsilon$. Furthermore, we may estimate the auxiliary epi-distance $\widehat{d}_{\rho+\varepsilon}^+(f_P, f_Q)$ from above by the uniform distance $d_{\mathcal{F}_\mathcal{U}}(P, Q)$ (cf. also Example 7.62 in Rockafellar and Wets (1998)).
It remains to note that the level sets $X_\varepsilon^*(P)$ and $X_\varepsilon^*(Q)$ are also bounded, since f_P and f_Q are lower semicontinuous and convex, and their solution sets are nonempty and bounded, respectively. Hence, we may choose the constant ρ large enough such that the equality $\mathbb{D}_\rho(X_\varepsilon^*(P), X_\varepsilon^*(Q)) = \mathbb{D}_\infty(X_\varepsilon^*(P), X_\varepsilon^*(Q))$ holds. This completes the proof. □

Most of the results in this and the previous section illuminate the role of the distance $d_{\mathcal{F}_\mathcal{U}}$ as a *minimal information (m.i.)* pseudometric for stability, i.e., as a pseudometric processing the minimal information of problem (1) and implying quantitative stability of its optimal values and solution sets. Furthermore, notice that all results remain valid when enlarging the set $\mathcal{F}_\mathcal{U}$ and, thus, bounding $d_{\mathcal{F}_\mathcal{U}}$ from above by another distance, and when reducing the set $\mathcal{P}_{\mathcal{F}_\mathcal{U}}$ to a subset on which such a distance is defined and finite.
Such a distance d_{id} bounding $d_{\mathcal{F}_\mathcal{U}}$ from above will be called an *ideal probability metric* associated with (1) if it has ζ-structure (9) generated by some class of functions $\mathcal{F} = \mathcal{F}_{\mathrm{id}}$ from Ξ to $\overline{\mathbb{R}}$ such that $\mathcal{F}_{\mathrm{id}}$ contains the functions $CF_j(x, \cdot)$ for each $x \in X \cap \mathrm{cl}\,\mathcal{U}$, $j = 0, \ldots, d$, and some normalizing constant $C > 0$, and such that any function in $\mathcal{F}_{\mathrm{id}}$ shares typical analytical properties with some function $F_j(x, \cdot)$.

In our applications of the general analysis in Section 3 we clarify such typical analytical properties. Here, we only mention that typical functions $F_j(x,\cdot)$ in stochastic programming are nondifferentiable, but piecewise locally Lipschitz continuous with discontinuities at boundaries of polyhedral sets. More precisely, function classes $\mathcal{F}$ contained in

$$\text{span}\,\{F\chi_B : F \in \mathcal{F}, B \in \mathcal{B}\}, \tag{27}$$

where $\mathcal{F} \subseteq \mathcal{F}_p(\Xi)$, $\mathcal{B} \subseteq \mathcal{B}_{\text{ph}_k}(\Xi)$ for some $p \geq 1$ and $k \in \mathbb{N}$, are candidates for an ideal class $\mathcal{F}_{\text{id}}$. The extremal cases, namely, $\mathcal{F}_p(\Xi)$ and $\mathcal{F}_{\mathcal{B}}$, are discussed in Section 2.1. To get an idea of how to associate an ideal metric with a stochastic program, we consider the p-th order Fortet-Mourier metric ζ_p introduced in Section 2.1. Then the following result is an immediate consequence of the general ones.

Corollary 14 *Let $d = 0$ and assume that*
(i) $X^*(P)$ is nonempty and $\mathcal{U}$ is an open, bounded neighbourhood of $X^*(P)$,
(ii) X is convex and $F_0(\cdot,\xi)$ is convex on $\mathbb{R}^m$ for each $\xi \in \Xi$,
(iii) there exist constants $L > 0$, $p \geq 1$ such that $\frac{1}{L}F_0(x,\cdot) \in \mathcal{F}_p(\Xi)$ for each $x \in X \cap \text{cl}\,\mathcal{U}$.
Then there exists a constant $\delta > 0$ such that

$$\begin{aligned} |\vartheta(P) - \vartheta(Q)| &\leq L\zeta_p(P,Q) \quad \text{and} \\ \emptyset \neq X^*(Q) &\subseteq X^*(P) + \Psi_P(L\zeta_p(P,Q))\mathbb{B} \end{aligned}$$

whenever $Q \in \mathcal{P}_p(\Xi)$ and $\zeta_p(P,Q) < \delta$. Here, the function Ψ_P is given by (23).

Proof: The assumptions of Theorem 5 are satisfied. Hence, the result is a consequence of the Theorems 5 and 9 and the fact that (iii) is equivalent to

$$|F_0(x,\xi) - F_0(x,\tilde{\xi})| \leq L\max\{1, \|\xi\|, \|\tilde{\xi}\|\}^{p-1}\|\xi - \tilde{\xi}\|$$

for each $\xi, \tilde{\xi} \in \Xi$ and $x \in X \cap \text{cl}\,\mathcal{U}$, and, thus, it implies $d_{\mathcal{F}_\mathcal{U}}(P,Q) \leq L\zeta_p(P,Q)$ for all $P, Q \in \mathcal{P}_p(\Xi)$. Furthermore, due to the convexity assumption (ii) the localized optimal values $\vartheta_\mathcal{U}$ and solution sets $X^*_\mathcal{U}$ may be replaced by ϑ and X^*, respectively, if Q is close to P (see Remark 11). □

Example 15 (newsboy continued)
In case of minimal expected costs the set $\mathcal{F}_\mathcal{U}$ is a specific class of piecewise linear functions of the form $\{(r-c)x + c\max\{0, x - \cdot\} : x \in X \cap \text{cl}\,\mathcal{U}\}$. Furthermore, $\int_\Xi F_0(x,\xi)dP(\xi)$ is also piecewise linear and Corollary 14 applies with $L := c$, $p := 1$ and a linear function Ψ_P. Hence, the solution set $X^*(\cdot)$

behaves upper Lipschitzian at $P \in \mathcal{P}_1(\mathbb{N})$ with respect to ζ_1, i.e.,

$$\sup_{x \in X^*(Q)} d(x, X^*(P)) \leq c\zeta_1(P,Q) = c \int_{\mathbb{R}} |F_P(r) - F_Q(r)| dr = c \sum_{k \in \mathbb{N}} \Big| \sum_{i=1}^{k} (\pi_i - \tilde{\pi}_i) \Big|.$$

Here, we made use of an explicit representation of the Kantorovich metric on $\mathcal{P}(\mathbb{R})$ (Section 5.4 in Rachev (1991)), and F_P and F_Q are the probability distribution functions of the measures $P = \sum_{k \in \mathbb{N}} \pi_k \delta_k$ and $Q = \sum_{k \in \mathbb{N}} \tilde{\pi}_k \delta_k$, respectively.

2.4 Mean-Risk Models

The expectation functional appearing in the basic model (1) is certainly not the only statistical parameter of interest of the (real-valued) cost or constraint functions F_j, $j = 0, \ldots, d$, with respect to P. *Risk functionals* or *risk measures* are regarded as statistical parameters of probability measures in $\mathcal{P}(\mathbb{R})$, i.e., they are mappings from subsets of $\mathcal{P}(\mathbb{R})$ to $\mathbb{R}$. When risk functionals are used in the context of the model (1), they are evaluated at the probability distributions $P[F_j(x, \cdot)]^{-1}$ for $x \in X$ and $j = 0, \ldots, d$. Practical risk management in decision making under uncertainty often requires to minimize or bound several risk functionals of the underlying distributions. Typical examples for risk functionals are (standard semi-) deviations, excess probabilities, value-at-risk, conditional value-at-risk etc. Some risk measures are defined as infima of certain (simple) stochastic optimization models (e.g. value-at-risk, conditional value-at-risk). Other measures are given as the expectation of a nonlinear function and, hence, their optimization fits into the framework of model (1) (e.g. expected utility functions, excess probabilities).
We refer to Section 4 of Pflug (2003) for an introduction to risk functionals and various examples, to Artzner et al. (1999), Delbaen (2002), Föllmer and Schied (2002) for a theory of coherent and convex risk measures, to Ogryczak and Ruszczyński (1999) for the relations to stochastic dominance and to Rockafellar and Uryasev (2002) for the role of the conditional value-at-risk.
Now, we assume that risk functionals $\mathbb{F}_j$, $j = 0, \ldots, d$ are given. In addition to the mean-risk model (2) we denote by Q a perturbation of the original probability measure P and consider the perturbed model

$$\min\{\mathbb{F}_0(Q[F_0(x, \cdot)]^{-1}) : x \in X, \mathbb{F}_j(Q[F_j(x, \cdot)]^{-1}) \leq 0, j = 1, ..., d\}. \tag{28}$$

To have all risk functionals $\mathbb{F}_j$ well defined, we assume for simplicity that they are given on the subset $\mathcal{P}_{\mathrm{b}}(\mathbb{R})$ of all probability measures in $\mathcal{P}(\mathbb{R})$ having bounded support. Then both models, (2) and (28), are well defined if we assume that all functions $F_j(x, \cdot)$ are bounded. Furthermore, we will need a continuity property of risk functionals.

A risk functional $\mathbb{F}$ on $\mathcal{P}_{\rm b}(\mathbb{R})$ is called *Lipschitz continuous w.r.t. to a class $\mathcal{H}$* of measurable functions from $\mathbb{R}$ to $\mathbb{R}$ if the estimate

$$|\mathbb{F}(G) - \mathbb{F}(\tilde{G})| \le \sup_{H \in \mathcal{H}} \Big| \int_{\mathbb{R}} H(r) d(G - \tilde{G})(r) \Big| \tag{29}$$

is valid for all $G, \tilde{G} \in \mathcal{P}_{\rm b}(\mathbb{R})$. The following examples and Proposition 8 in Pflug (2003) show that many risk functionals satisfy such a Lipschitz property.

Example 16 We consider the *conditional value-at-risk* of a probability distribution $G \in \mathcal{P}_{\rm b}(\mathbb{R})$ at level $p \in (0,1)$, which is defined by

$$\mathrm{CVaR}_p(G) := \inf \Big\{ r + \frac{1}{1-p} \int_{\mathbb{R}} \max\{0, \xi - r\} dG(\xi) : r \in \mathbb{R} \Big\}.$$

Hence, $\mathrm{CVaR}_p(G)$ is the optimal value of a stochastic program with recourse (see Section 3.1). Clearly, the estimate

$$|\mathrm{CVaR}_p(G) - \mathrm{CVaR}_p(\tilde{G})| \le \frac{1}{1-p} \sup_{r \in \mathbb{R}} \Big| \int_{\mathbb{R}} \max\{0, \xi - r\} d(G - \tilde{G})(\xi) \Big|$$

is valid for all $G, \tilde{G} \in \mathcal{P}_{\rm b}(\mathbb{R})$. Hence, the conditional value-at-risk is Lipschitz continuous w.r.t. the class $\mathcal{H} := \{\max\{0, \cdot - r\} : r \in \mathbb{R}\}$.
The *value-at-risk* of $G \in \mathcal{P}_{\rm b}(\mathbb{R})$ at level $p \in (0,1)$ is given by

$$\mathrm{VaR}_p(G) := \inf \{ r \in \mathbb{R} : G(\xi \le r) \ge p \}.$$

Thus, $\mathrm{VaR}_p(G)$ is the optimal value of a chance constrained stochastic program. In Section 3.3 it is shown that the metric regularity of the mapping $r \mapsto \{y \in \mathbb{R} : G(\xi \le r) \ge p - y\}$ at pairs $(\bar{r}, 0)$ with $\bar{r} \in X^*(G)$ is indispenable for Lipschitz continuity properties of the optimal value. If the metric regularity property is satisfied for the measure G and the level p, we obtain, from Theorem 39, the estimate

$$|\mathrm{VaR}_p(G) - \mathrm{VaR}_p(\tilde{G})| \le L d_{\rm K}(G, \tilde{G}) = \sup_{r \in \mathbb{R}} \Big| \int_{\mathbb{R}} L \chi_{(-\infty, r]}(\xi) d(G - \tilde{G})(\xi) \Big|$$

for some constant $L > 0$ and sufficiently small Kolmogorov distance $d_{\rm K}(G, \tilde{G})$. Hence, the corresponding class of functions is $\mathcal{H} := \{L \chi_{(-\infty, r]} : r \in \mathbb{R}\}$. We note that the metric regularity requirement may lead to serious complications when using the value-at-risk in stochastic programming models because $\mathrm{VaR}_p(\cdot)$ has to be evaluated at measures depending on x.

Example 17 The *upper semi-deviation* $sd_+(G)$ of a measure $G \in \mathcal{P}_{\rm b}(\mathbb{R})$, which is defined by

$$sd_+(G) := \int_{\mathbb{R}} \max\Big\{0, \xi - \int_{\mathbb{R}} u dG(u)\Big\} dG(\xi),$$

is Lipschitz continuous w.r.t. the class $\mathcal{H} := \{\max\{0, \cdot - r\} + \cdot : r \in \mathbb{R}\}$.

The examples indicate that typical Lipschitz continuity classes $\mathcal{H}$ of risk functionals contain products of some functions in $\mathcal{F}_k(\mathbb{R})$ for some $k \in \mathbb{N}$ and of characteristic functions $\chi_{(-\infty,r]}$ for some $r \in \mathbb{R}$. Hence, their structure is strongly related to that of the ideal function classes (27) for stability.
To state our main stability result for the model (2), let $\mathcal{X}(P)$, $\vartheta(P)$, $X^*(P)$ denote the following more general quantities in this section:

$$\begin{aligned} \mathcal{X}(P) &:= \{x \in X : \mathbb{F}_j(P[F_j(x,\cdot)]^{-1}) \leq 0, j = 1, ..., d\}, \\ \vartheta(P) &:= \inf\{\mathbb{F}_0(P[F_0(x,\cdot)]^{-1}) : x \in \mathcal{X}(P)\}, \\ X^*(P) &:= \{x \in \mathcal{X}(P) : \mathbb{F}_0(P[F_0(x,\cdot)]^{-1}) = \vartheta(P)\}. \end{aligned}$$

The localized notions $\vartheta_{\mathcal{U}}(P)$ and $X^*_{\mathcal{U}}(P)$ are defined accordingly.

Theorem 18 *For each $j = 0, \ldots, d$, let the function F_j be uniformly bounded and the risk functional $\mathbb{F}_j$ be Lipschitz continuous on $\mathcal{P}_{\rm b}(\mathbb{R})$ w.r.t. some class $\mathcal{H}_j$ of measurable functions from $\mathbb{R}$ to $\mathbb{R}$. Let $P \in \mathcal{P}(\Xi)$ and assume that*

(i) *$X^*(P) \neq \emptyset$ and $\mathcal{U} \subseteq \mathbb{R}^m$ is an open bounded neighbourhood of $X^*(P)$,*
(ii) *if $d \geq 1$, the function $x \mapsto \mathbb{F}_0(P[F_0(x,\cdot)]^{-1})$ is Lipschitz continuous on $X \cap \mathrm{cl}\,\mathcal{U}$,*
(iii) *the mapping $x \mapsto \{y \in \mathbb{R}^d : x \in X, \mathbb{F}_j(P[F_j(x,\cdot)]^{-1}) \leq y_j, j = 1, ..., d\}$ from $\mathbb{R}^m$ to $\mathbb{R}^d$ is metrically regular at each pair $(\bar{x}, 0)$ with $\bar{x} \in X^*(P)$.*

Then there exist constants $L > 0$ and $\delta > 0$ such that the estimates

$$\begin{aligned} |\vartheta(P) - \vartheta_{\mathcal{U}}(Q)| &\leq L d_{\mathcal{F}^{\mathcal{H}}_{\mathcal{U}}}(P, Q) \\ \emptyset \neq X^*_{\mathcal{U}}(Q) &\subseteq X^*(P) + L\Psi_P(d_{\mathcal{F}^{\mathcal{H}}_{\mathcal{U}}}(P, Q))\mathbb{B} \end{aligned}$$

are valid whenever $Q \in \mathcal{P}(\Xi)$ and $d_{\mathcal{F}^{\mathcal{H}}_{\mathcal{U}}}(P, Q) < \delta$. Here, Ψ_P is given by (23) and the distance $d_{\mathcal{F}^{\mathcal{H}}_{\mathcal{U}}}$ is defined by

$$d_{\mathcal{F}^{\mathcal{H}}_{\mathcal{U}}}(P, Q) := \sup_{\substack{j=0,\ldots,d \\ x \in X \cap \mathrm{cl}\,\mathcal{U} \\ H_j \in \mathcal{H}_j}} \Big| \int_{\Xi} H_j(F_j(x,\xi))(P - Q)(d\xi) \Big|.$$

Proof: We proceed as in the proofs of Theorems 5 and 9, but now we use the distance

$$\hat{d}_{\mathbb{F}}(P,Q) := \sup_{\substack{j=0,\ldots,d\\ x\in X\cap \mathrm{cl}\,\mathcal{U}}} \left|\mathbb{F}_j\left(P[F_j(x,\cdot)]^{-1}\right) - \mathbb{F}_j\left(Q[F_j(x,\cdot)]^{-1}\right)\right|$$

instead of $d_{\mathcal{F}_\mathcal{U}}$. In this way we obtain constants $L > 0$, $\delta > 0$ and the estimates

$$|\vartheta(P) - \vartheta_\mathcal{U}(Q)| \le L\hat{d}_{\mathbb{F}}(P,Q)$$
$$\emptyset \ne X^*_\mathcal{U}(Q) \subseteq X^*(P) + L\Psi_P(\hat{d}_{\mathbb{F}}(P,Q))\mathbb{B}$$

for each $Q \in \mathcal{P}(\Xi)$ such that $\hat{d}_{\mathbb{F}}(P,Q) < \delta$. It remains to appeal to the estimate

$$\hat{d}_{\mathbb{F}}(P,Q) \le \sup_{\substack{j=0,\ldots,d\\ x\in X\cap \mathrm{cl}\,\mathcal{U}}} \sup_{H_j\in\mathcal{H}_j} \Big|\int_{\mathbb{R}} H_j(r)d((P-Q)[F_j(x,\cdot)]^{-1})(r)\Big| = d_{\mathcal{F}^{\mathcal{H}}_{\mathcal{U}}}(P,Q),$$

which is a consequence of the Lipschitz continuity (29) of the risk functionals $\mathbb{F}_j$, $j = 0,\ldots,d$. $\square$

The result implies that stability properties of the mean-risk model (2) containing risk functionals $\mathbb{F}_j$ with Lipschitz continuity classes $\mathcal{H}_j$, $j = 0,\ldots,d$, depend on the class

$$\mathcal{F}^{\mathcal{H}}_{\mathcal{U}} := \{H_j(F_j(x,\cdot)) : x \in X \cap \mathrm{cl}\,\mathcal{U}, H_j \in \mathcal{H}_j, j = 0,\ldots,d\}$$

instead of $\mathcal{F}_\mathcal{U}$ in case of model (1). Hence, the stability behaviour may change considerably when replacing the expectation functionals in (1) by other risk functionals. For example, the newsboy model based on minimal expected costs behaves stable at all $P \in \mathcal{P}_1(\mathbb{N})$ (Example 15), but the minimum risk variant of the model (see Example 1) may become unstable.

Example 19 (newsboy continued)
We consider the chance constrained model (3) whose solution set is $X^*(P) = \{(k,0)\}$ with the maximal k such that $\sum_{i=k}^{\infty}\pi_i \ge p$ in its first component. We assume that equality $\sum_{i=k}^{\infty}\pi_i = p$ and $\pi_k > 0$ holds. To establish instability, we consider the approximations $P_n := \sum_{i=1}^{\infty}\pi_i^{(n)}\delta_i$ of P, where $\pi_i^{(n)} := \pi_i$ for all $i \notin \{k-1,k\}$ and $\pi_{k-1}^{(n)} := \pi_{k-1} + \frac{1}{n}$, $\pi_k^{(n)} := \pi_k - \frac{1}{n}$ for sufficiently large $n \in \mathbb{N}$ such that $\pi_k - \frac{1}{n} > 0$. Then the perturbed solution set is $X^*(P_n) = \{(k-1,0)\}$ for any sufficiently large n. On the other hand, we obtain for the Kolmogorov distance $d_K(P,P_n) = \frac{1}{n}$, i.e., weak convergence of (P_n) to P. Furthermore, the model (3) is stable with respect to the metric d_K at each $P = \sum_{i=1}^{\infty}\pi_i\delta_i \in \mathcal{P}(\mathbb{N})$ such that $\sum_{i=1}^{k}\pi_i \ne 1-p$ for each $k \in \mathbb{N}$. The latter fact is a consequence of

Theorem 5 as the metric regularity condition is satisfied (see also Remark 2.5 in Römisch and Schultz (1991b)).

However, if the conditional value-at-risk or the upper semi-deviation are incorporated into the objective of (mixed-integer) two-stage stochastic programs, their ideal function classes and, thus, their ideal metrics (see Sections 3.1 and 3.2) do not change. These observations are immediate consequences of the following more general conclusion of the previous theorem.

Corollary 20 *Let $d = 0$. We consider the stochastic programming model*

$$\min\{\mathbb{F}_0(P[F_0(x,\cdot)]^{-1}) : x \in X\}, \tag{30}$$

where F_0 is uniformly bounded and the risk functional $\mathbb{F}_0$ is Lipschitz continuous on $\mathcal{P}_b(\mathbb{R})$ w.r.t. some class $\mathcal{H}_0$.
Let $P \in \mathcal{P}(\Xi)$, $X^*(P) \neq \emptyset$ and $\mathcal{U}$ be an open bounded neighbourhood of $X^*(P)$. Assume that $\{F_0(x,\cdot) : x \in X \cap \operatorname{cl}\mathcal{U}\}$ is contained in some class $\mathcal{F}_c$ of functions from Ξ to $\mathbb{R}$ and $H \circ F \in L_0\mathcal{F}_c$ holds for all $H \in \mathcal{H}_0$, $F \in \mathcal{F}_c$ and some positive constant L_0.
Then there exist constants $L > 0$ and $\delta > 0$ such that the estimates

$$|\vartheta(P) - \vartheta_{\mathcal{U}}(Q)| \leq L d_{\mathcal{F}_c}(P,Q)$$
$$\emptyset \neq X^*_{\mathcal{U}}(Q) \subseteq X^*(P) + L\Psi_P(d_{\mathcal{F}_c}(P,Q))\mathbb{B}$$

are valid whenever $Q \in \mathcal{P}(\Xi)$ and $d_{\mathcal{F}_c}(P,Q) < \delta$.

Proof: Clearly, we have in that case $d_{\mathcal{F}^{\mathcal{H}}_{\mathcal{U}}}(P,Q) \leq L_0 d_{\mathcal{F}_c}(P,Q)$. □

Important examples for $\mathcal{H}_0$ and $\mathcal{F}_c$ are multiples of $\mathcal{F}_1(\mathbb{R})$ and of $\mathcal{F}_p(\Xi)$ (for $p \geq 1$) and $\{F\chi_B : F \in \mathcal{F}_1(\Xi), B \in \mathcal{B}\}$, respectively.

3 Stability of Two-Stage and Chance Constrained Programs

3.1 Linear Two-Stage Models

We consider the linear two-stage stochastic program with fixed recourse

$$\min\Big\{\langle c,x\rangle + \int_{\Xi}\langle q(\xi),y(\xi)\rangle dP(\xi) : Wy(\xi) = h(\xi) - T(\xi)x, \qquad y(\xi) \geq 0,\ x \in X\Big\}, \tag{31}$$

where $c \in \mathbb{R}^m$, $X \subseteq \mathbb{R}^m$ and $\Xi \subseteq \mathbb{R}^s$ are convex polyhedral, W is an $(r,\overline{m})$-matrix, $P \in \mathcal{P}(\Xi)$, and the vectors $q(\xi) \in \mathbb{R}^{\overline{m}}$, $h(\xi) \in \mathbb{R}^r$ and the (r,m)-matrix $T(\xi)$ depend affine linearly on $\xi \in \Xi$. The latter assumption covers many practical situations. At the same time, it avoids the inclusion of all components of the recourse costs, the technology matrix and the right-hand side into ξ, because this could lead to serious restrictions when imposing additional conditions on P. We define the function $F_0 : \mathbb{R}^m \times \Xi \to \overline{\mathbb{R}}$ by

$$F_0(x,\xi) = \begin{cases} \langle c,x\rangle + \Phi(q(\xi),h(\xi)-T(\xi)x) & , h(\xi)-T(\xi)x \in \operatorname{pos} W \, , q(\xi) \in D \\ +\infty & , \text{otherwise} \end{cases}$$

where $\operatorname{pos} W = \{Wy : y \in \mathbb{R}^{\overline{m}}_+\}$, $D = \{u \in \mathbb{R}^{\overline{m}} : \{z \in \mathbb{R}^r : W'z \leq u\} \neq \emptyset\}$ (with W' denoting the transpose of the matrix W) and $\Phi(u,t) = \inf\{\langle u,y\rangle : Wy = t, y \geq 0\}$ $\quad ((u,t) \in \mathbb{R}^{\overline{m}} \times \mathbb{R}^r)$. Then problem (31) may be rewritten equivalently as a minimization problem with respect to the first stage decision x, namely,

$$\min \Big\{ \int_\Xi F_0(x,\xi) dP(\xi) : \; x \in X \Big\}. \tag{32}$$

In order to utilize the general stability results of Section 2, we need a characterization of the continuity and growth properties of the function F_0. As a first step we recall some well-known properties of the function Φ, which were derived in Walkup and Wets (1969a).

Lemma 21 *The function Φ is finite and continuous on the $(\overline{m}+r)$-dimensional polyhedral cone $D \times \operatorname{pos} W$ and there exist $(r,\overline{m})$-matrices C_j and $(\overline{m}+r)$-dimensional polyhedral cones $\mathcal{K}_j$, j=1,...,N, such that*

$$\bigcup_{j=1}^{N} \mathcal{K}_j = D \times \operatorname{pos} W \; , \; \operatorname{int} \mathcal{K}_i \cap \operatorname{int} \mathcal{K}_j = \emptyset \, , \; i \neq j,$$
$$\Phi(u,t) = \langle C_j u, t\rangle, \quad \text{for each} \;\; (u,t) \in \mathcal{K}_j, \; j = 1,...,N.$$

Moreover, $\Phi(u,\cdot)$ is convex on $\operatorname{pos} W$ *for each $u \in D$, and $\Phi(\cdot,t)$ is concave on D for each $t \in$* $\operatorname{pos} W$.

To have problem (32) well defined we introduce the following assumptions:
(A1) For each $(x,\xi) \in X \times \Xi$ it holds that $h(\xi) - T(\xi)x \in \operatorname{pos} W$ and $q(\xi) \in D$.
(A2) $P \in \mathcal{P}_2(\Xi)$, i.e., $\int_\Xi \|\xi\|^2 dP(\xi) < \infty$.
Condition (A1) sheds some light on the role of the set Ξ. Due to the affine linearity of $q(\cdot)$, $h(\cdot)$ and $T(\cdot)$ the polyhedrality assumption on Ξ is not restrictive. (A1) combines the two usual conditions: *relatively complete recourse*

and *dual feasibility*. It implies that $X \times \Xi \subseteq \operatorname{dom} F_0$.

Proposition 22 *Let (A1) be satisfied. Then F_0 is a random convex function. Furthermore, there exist constants $L > 0$, $\hat{L} > 0$ and $K > 0$ such that the following holds for all $\xi, \tilde{\xi} \in \Xi$ and $x, \tilde{x} \in X$ with $\max\{\|x\|, \|\tilde{x}\|\} \leq r$:*

$$\begin{aligned} |F_0(x,\xi) - F_0(x,\tilde{\xi})| &\leq Lr \max\{1, \|\xi\|, \|\tilde{\xi}\|\} \|\xi - \tilde{\xi}\|, \\ |F_0(x,\xi) - F_0(\tilde{x},\xi)| &\leq \hat{L} \max\{1, \|\xi\|^2\} \|x - \tilde{x}\|, \\ |F_0(x,\xi)| &\leq Kr \max\{1, \|\xi\|^2\}. \end{aligned}$$

Proof: From Lemma 21 and (A1) we conclude that F_0 is continuous on $\operatorname{dom} F_0$ and, hence, on $X \times \Xi$. This implies that F_0 is a random lower semicontinuous function (cf. Example 14.31 in Rockafellar and Wets (1998)). It is a random convex function since the properties of Φ in Lemma 21 imply that $F_0(\cdot, \xi)$ is convex for each $\xi \in \Xi$. In order to verify the Lipschitz property of F_0, let $x \in X$ with $\|x\| \leq r$ and consider, for each $j = 1, \ldots, N$, and $\xi \in \Xi_j$ the function

$$g_j(\xi) := F_0(x,\xi) = \Phi(q(\xi), h(\xi) - T(\xi)x) = \langle C_j q(\xi), h(\xi) - T(\xi)x \rangle,$$

where the sets $\Xi_j := \{\xi \in \Xi : (q(\xi), h(\xi) - T(\xi)x) \in \mathcal{K}_j\}$ are polyhedral, and C_j and $\mathcal{K}_j$ are the matrices and the polyhedral cones from Lemma 21, respectively. Since $q(\cdot)$, $h(\cdot)$ and $T(\cdot)$ depend affine linearly on ξ, the function g_j depends quadratically on ξ and linearly on x. Hence, there exists a constant $L_j > 0$ such that g_j satisfies the following Lipschitz property:

$$|g_j(\xi) - g_j(\tilde{\xi})| \leq L_j r \max\{1, \|\xi\|, \|\tilde{\xi}\|\} \|\xi - \tilde{\xi}\| \quad \text{for all } \xi, \tilde{\xi} \in \Xi_j.$$

Now, let $\xi, \tilde{\xi} \in \Xi$, assume that $\xi \in \Xi_i$ and $\tilde{\xi} \in \Xi_k$ for some $i, k \in \{1, \ldots, N\}$ and consider the line segment $[\xi, \tilde{\xi}] = \{\xi(\lambda) = (1 - \lambda)\xi + \lambda\tilde{\xi} : \lambda \in [0, 1]\}$. Since $[\xi, \tilde{\xi}] \subseteq \Xi$, there exist indices $i_j, j = 1, \ldots, l$, such that $i_1 = i$, $i_l = k$, $[\xi, \tilde{\xi}] \cap \Xi_{i_j} \neq \emptyset$ for each $j = 1, \ldots, l$ and $[\xi, \tilde{\xi}] \subseteq \bigcup_{j=1}^{l} \Xi_{i_j}$. Furthermore, there exist increasing numbers $\lambda_{i_j} \in [0, 1]$ for $j = 0, \ldots, l - 1$ such that $\xi(\lambda_{i_0}) = \xi(0) = \xi$, $\xi(\lambda_{i_j}) \in \Xi_{i_j} \cap \Xi_{i_{j+1}}$ and $\xi(\lambda) \notin \Xi_{i_j}$ if $\lambda_{i_j} < \lambda \leq 1$. Then we obtain

$$\begin{aligned} |F_0(x,\xi) - F_0(x,\tilde{\xi})| &= |g_{i_1}(\xi) - g_{i_l}(\tilde{\xi})| \\ &\leq \sum_{j=0}^{l-1} |g_{i_{j+1}}(\xi(\lambda_{i_j})) - g_{i_{j+1}}(\xi(\lambda_{i_{j+1}}))| \\ &\leq \sum_{j=0}^{l-1} L_{i_{j+1}} r \max\{1, \|\xi\|, \|\tilde{\xi}\|\} \|\xi(\lambda_{i_j}) - \xi(\lambda_{i_{j+1}})\| \end{aligned}$$

$$\leq \max_{j=1,\dots,N} L_j r \max\{1, \|\xi\|, \|\tilde{\xi}\|\} \sum_{j=0}^{l-1} \|\xi(\lambda_{i_j}) - \xi(\lambda_{i_{j+1}})\|$$
$$\leq \max_{j=1,\dots,N} L_j r \max\{1, \|\xi\|, \|\tilde{\xi}\|\} \|\xi - \tilde{\xi}\|,$$

where we have used for the last three estimates that $\|\xi(\lambda)\| \leq \max\{\|\xi\|, \|\tilde{\xi}\|\}$ for each $\lambda \in [0,1]$ and $|\lambda - \tilde{\lambda}|\|\xi - \tilde{\xi}\| = \|\xi(\lambda) - \xi(\tilde{\lambda})\|$ holds for all $\lambda, \tilde{\lambda} \in [0,1]$. Lipschitz continuity of F_0 with respect to x is shown in Theorem 10 of Kall (1976) and in Theorem 7.7 of Wets (1974). In particular, the second estimate of the proposition is a consequence of those results. Furthermore, from Lemma 21 we conclude the estimate

$$|F_0(x,\xi)| \leq \sup_{\|x\| \leq r} \{|\langle c, x\rangle| + \max_{j=1,\dots,N} |\langle C_j q(\xi), h(\xi) - T(\xi)x\rangle|\}$$
$$\leq \|c\| r + (\max_{j=1,\dots,N} \|C_j\|) \|q(\xi)\| (\|h(\xi)\| + \|T(\xi)\| r)$$

for any pair $(x,\xi) \in X \times \Xi$ with $\|x\| \leq r$. Then the third estimate follows again from the fact that $q(\cdot)$, $h(\cdot)$ and $T(\cdot)$ depend affine linearly on ξ. □

The estimate in Proposition 22 implies that, for any $r > 0$, any nonempty bounded $\mathcal{U} \subseteq \mathbb{R}^m$ and some $\rho > 0$, it holds that

$$\int_\Xi \inf_{\substack{x \in X \\ \|x\| \leq r}} F_0(x,\xi) dQ(\xi) \geq -Kr(1 + \int_\Xi \|\xi\|^2 dQ(\xi)) > -\infty\,,$$
$$\sup_{x \in X \cap \mathcal{U}} |\int_\Xi F_0(x,\xi) dQ(\xi)| \leq K\rho(1 + \int_\Xi \|\xi\|^2 dQ(\xi)) < \infty\,,$$

if $Q \in \mathcal{P}(\Xi)$ has a finite second order moment. Hence, for any nonempty bounded $\mathcal{U} \subseteq \mathbb{R}^m$ the set of probability measures $\mathcal{P}_{\mathcal{F}_\mathcal{U}}$ contains the set of measures on Ξ having finite second order moments, i.e.,

$$\mathcal{P}_{\mathcal{F}_\mathcal{U}} \supseteq \left\{Q \in \mathcal{P}(\Xi) : \int_\Xi \|\xi\|^2 dQ(\xi) < \infty\right\} = \mathcal{P}_2(\Xi).$$

The following stability results for optimal values and solution sets of the two-stage problem (32) are now a direct consequence of the results of Section 2.

Theorem 23 *Let (A1) and (A2) be satisfied and let $X^*(P)$ be nonempty and $\mathcal{U}$ be an open, bounded neighbourhood of $X^*(P)$.*
Then there exist constants $L > 0$ and $\delta > 0$ such that

$$|\vartheta(P) - \vartheta(Q)| \leq L\zeta_2(P,Q) \quad \text{and}$$
$$\emptyset \neq X^*(Q) \subseteq X^*(P) + \Psi_P(L\zeta_2(P,Q))\mathbb{B}$$

whenever $Q \in \mathcal{P}_2(\Xi)$ *and* $\zeta_2(P,Q) < \delta$*, where* Ψ_P *is given by (23).*

Proof: The result is a consequence of Corollary 14 with $p = 2$. The assumptions (ii) and (iii) of Corollary 14 are verified in Proposition 22. □

Theorem 24 *Let (A1) and (A2) be satisfied and let* $X^*(P)$ *be nonempty and bounded. Then there exist constants* $\bar{L} > 0$ *and* $\bar{\varepsilon} > 0$ *such that the estimate*

$$\mathbb{D}_\infty(X^*_\varepsilon(P), X^*_\varepsilon(Q)) \leq \frac{\bar{L}}{\varepsilon}\zeta_2(P,Q)$$

holds for any $\varepsilon \in (0,\bar{\varepsilon})$ *and* $Q \in \mathcal{P}_2(\Xi)$ *such that* $\zeta_2(P,Q) < \varepsilon$*. Here,* $\mathbb{D}_\infty$ *denotes the Pompeiu-Hausdorff distance (26).*

Proof: Since the assumptions of Theorem 13 are satisfied, we conclude that there exist constants $\rho > 0$ and $\bar{\varepsilon} > 0$ such that

$$\mathbb{D}_\infty(X^*_\varepsilon(P), X^*_\varepsilon(Q)) \leq \frac{2\rho}{\varepsilon} d_{\mathcal{F}_\mathcal{U}}(P,Q)$$

holds for $\mathcal{U} := (\rho + \bar{\varepsilon})\mathbb{B}$ and any $\varepsilon \in (0,\bar{\varepsilon})$, $Q \in \mathcal{P}_{\mathcal{F}_\mathcal{U}}$ such that $d_{\mathcal{F}_\mathcal{U}}(P,Q) < \varepsilon$. Proposition 22 implies the estimate $d_{\mathcal{F}_\mathcal{U}}(P,Q) \leq L(\rho + \bar{\varepsilon})\zeta_2(P,Q)$, for some constant $L > 0$, which completes the proof. □

The theorems establish the quantitative stability of $\vartheta(\cdot)$ and $X^*(\cdot)$ and the Lipschitz stability of $X^*_\varepsilon(\cdot)$ with respect to ζ_2 in case of two-stage models with fixed recourse for fairly general situations. In case that either only the recourse costs or only the technology matrix and right-hand side are random, both results are valid for $(\mathcal{P}_1(\Xi), \zeta_1)$ instead of $(\mathcal{P}_2(\Xi), \zeta_2)$. We verify this observation for the corresponding conclusion of Theorem 23.

Corollary 25 *Let either only* $q(\cdot)$ *or only* $T(\cdot)$ *and* $h(\cdot)$ *be random and (A1) be satisfied. Let* $P \in \mathcal{P}_1(\Xi)$*,* $X^*(P)$ *be nonempty and* $\mathcal{U}$ *be an open, bounded neighbourhood of* $X^*(P)$*. Then there exist constants* $L > 0$*,* $\delta > 0$ *such that*

$$\begin{aligned} |\vartheta(P) - \vartheta(Q)| &\leq L\zeta_1(P,Q) \quad \text{and} \\ \emptyset \neq X^*(Q) &\subseteq X^*(P) + \Psi_P(L\zeta_1(P,Q))\mathbb{B} \end{aligned}$$

whenever $Q \in \mathcal{P}_1(\Xi)$ *and* $\zeta_1(P,Q) < \delta$*, where* Ψ_P *is given by (23).*

Proof: By inspecting the proof of Proposition 22 one observes that now the function F_0 satisfies the following continuity and growth properties for all $\xi, \tilde{\xi} \in \Xi$ and $x, \tilde{x} \in X$ with $\max\{\|x\|, \|\tilde{x}\|\} \leq r$:

$$|F_0(x,\xi) - F_0(x,\tilde{\xi})| \leq Lr\|\xi - \tilde{\xi}\|,$$

$$|F_0(x,\xi)| \le Kr\max\{1, \|\xi\|\}.$$

Hence, the set $\mathcal{P}_{\mathcal{F}_\mathcal{U}}$ contains $\mathcal{P}_1(\Xi)$ and Corollary 14 applies with $p = 1$. $\square$

Next we provide some examples of recourse models showing that, in general, the estimate for solution sets in Theorem 23 is the best possible one and that $X^*(\cdot)$ is not lower semicontinuous at P if $X^*(P)$ is not a singleton.
All examples exploit the specific structure provided by the *simple recourse* condition, i.e., $\overline{m} = 2s$, $q = (q_+, q_-)$ and $W = (I, -I)$, where $q_+, q_- \in \mathbb{R}^s$ and I is the (s,s)-identity matrix. Then $\operatorname{pos} W = \mathbb{R}^s$ holds and, hence, (A1) is satisfied iff $q \in D$, which is equivalent to the condition $q_+ + q_- \ge 0$, and

$$\Phi(q,t) = \sup\{\langle t, u\rangle : -q_- \le u \le q_+\}.$$

Example 26 Let $m = s = r = 1$, $\overline{m} = 2$, $c = 0$, $W = (1,-1)$, $X = [-1,1]$, $\Xi = \mathbb{R}$, $q(\xi) = (1,1)$, $T(\xi) = 1$, $h(\xi) = \xi$, $\forall \xi \in \Xi$. Let $P \in \mathcal{P}(\mathbb{R})$ be the uniform distribution on the interval $[-\frac{1}{2}, \frac{1}{2}]$. Then $\vartheta(P) = 1$, $X^*(P) = \{0\}$, and quadratic growth

$$\int_\Xi F_0(x,\xi)dP(\xi) = \int_{-\frac{1}{2}}^{\frac{1}{2}} |\xi - x| d\xi = \frac{1}{4} + x^2 = \vartheta(P) + d(x, X^*(P))^2$$

holds for each $x \in [-\frac{1}{2}, \frac{1}{2}]$. Let us consider the following perturbations $P_n \in \mathcal{P}(\mathbb{R})$ of P for $n > 4$ given by

$$P_n = (\frac{1}{2} - \varepsilon_n)(P_{ln} + P_{rn}) + \varepsilon_n(\delta_{-\varepsilon_n} + \delta_{\varepsilon_n}),$$

where $\varepsilon = n^{-\frac{1}{2}}$, P_{ln} and P_{rn} are the uniform distributions on $[-\frac{1}{2}, -\varepsilon_n)$ and $(\varepsilon_n, \frac{1}{2}]$, respectively, and δ_r is the measure placing unit mass at r. Using the explicit representation of ζ_1 in case of probability distributions on $\mathbb{R}$ (see Chapter 5.4 of Rachev (1991)), we obtain

$$\zeta_1(P, P_n) = \int_{-\infty}^{\infty} |P((-\infty,\xi]) - P_n((-\infty,\xi])| d\xi = \frac{1}{n} = \varepsilon_n^2.$$

Furthermore, it holds that $\vartheta(P_n) = \frac{1}{2}(\varepsilon_n^2 + \frac{1}{4})$, $X^*(P_n) = [-\varepsilon_n, \varepsilon_n]$ and, hence, $|\vartheta(P) - \vartheta(P_n)| = \frac{1}{2}\varepsilon_n^2$ and $\sup_{x\in X^*(P_n)} d(x, X^*(P)) = \varepsilon_n$ for each $n \in \mathbb{N}$. Hence, the estimate in Theorem 23 is best possible.
Next we consider the distribution $\hat{P} = \frac{1}{2}(\delta_{-\frac{1}{2}} + \delta_{\frac{1}{2}})$. Then we have $\vartheta(\hat{P}) = \frac{1}{2}$ and $X^*(\hat{P}) = [-\frac{1}{2}, \frac{1}{2}]$ and the linear growth condition

$$\int_{\Xi} F_0(x,\xi)d\hat{P}(\xi) = \int_{\Xi} |\xi - x| d\hat{P}(\xi) = \frac{1}{2}(|x + \frac{1}{2}| + |x - \frac{1}{2}|)$$
$$\geq \vartheta(\hat{P}) + d(x, X^*(\hat{P}))$$

for each $x \in X$. Consider the perturbations $\hat{P}_n = (1 - \frac{1}{n})\hat{P} + \frac{1}{n}\delta_0$ $(n \in \mathbb{N})$ of $\hat{P}$. Then

$$\zeta_1(\hat{P}, \hat{P}_n) = \int_{-\infty}^{\infty} |\hat{P}((-\infty, \xi]) - \hat{P}_n((-\infty, \xi])| d\xi = \frac{1}{2n},$$

holds for each $n \in \mathbb{N}$, where we have again used the explicit representation of ζ_1 in case of probability measures on $\mathbb{R}$. Furthermore, it holds that $\vartheta(\hat{P}_n) = (1 - \frac{1}{n})\frac{1}{2}$ and $X^*(\hat{P}_n) = \{0\}$ for each $n \in \mathbb{N}$. Hence, we have $\sup_{x \in X^*(\hat{P})} d(x, X^*(\hat{P}_n)) = \frac{1}{2}$.

Next we consider models with a stochastic technology matrix and recourse costs, respectively, and show that in such cases $X^*(\cdot)$ is also not lower semicontinuous at P, in general.

Example 27 Let $m = s = r = 1$, $\overline{m} = 2$, $c = 0$, $W = (1, -1)$, $X = [0, 1]$, $\Xi = \mathbb{R}_+$, $h(\xi) = 0$, $\forall \xi \in \Xi$.
In the first case, we set $q(\xi) = (1, 1)$ and $T(\xi) = -\xi$, $\forall \xi \in \Xi$.
In the second case, we set $q(\xi) = (\xi, \xi)$ and $T(\xi) = -1$, $\forall \xi \in \Xi$.
In both cases (A1) is satisfied. We consider $P = \delta_0$ and $P_n = \delta_{\frac{1}{n}}$, i.e., the unit masses at 0 and $\frac{1}{n}$, respectively, for each $n \in \mathbb{N}$. Clearly, (P_n) converges with respect to the metric ζ_1 to P in $\mathcal{P}_1(\mathbb{R})$. Furthermore, in both cases

$$\int_{\Xi} F_0(x,\xi)dP_n(\xi) = \int_{\Xi} \xi x \, dP_n(\xi) = \frac{x}{n}$$

holds for each $x \in X$. Then $X^*(P) = X$ and $X^*(P_n) = \{0\}$ for any $n \in \mathbb{N}$, which implies $\sup_{x \in X^*(P)} d(x, X^*(P_n)) = 1$.

The examples show that continuity properties of $X^*(\cdot)$ at P in terms of the Pompeiu-Hausdorff distance cannot be achieved in general unless $X^*(P)$ is a singleton. Nevertheless, we finally establish such quantitative stability results for models where the technology matrix is fixed, i.e., $T(\xi) \equiv T$, and a specific nonuniqueness of $X^*(P)$ is admitted. For their derivation we need an argument that decomposes the original two-stage stochastic program into another two-stage program with decisions taken from $T(X)$ and a parametric linear program not depending on P.

Lemma 28 *Let (A1) be satisfied and let $Q \in \mathcal{P}_2(\Xi)$ be such that $X^*(Q)$ is nonempty. Then we have*

$$\begin{aligned}
\vartheta(Q) &= \inf\Big\{\pi(\chi) + \int_{\Xi} \Phi(q(\xi), h(\xi) - \chi)dQ(\xi) : \chi \in T(X)\Big\} \\
&= \pi(Tx) + \int_{\Xi} \Phi(q(\xi), h(\xi) - Tx)dQ(\xi), \ \forall x \in X^*(Q), \\
X^*(Q) &= \sigma(Y^*(Q)), \qquad \textit{where} \\
Y^*(Q) &:= \text{argmin}\Big\{\pi(\chi) + \int_{\Xi} \Phi(q(\xi), h(\xi) - \chi)dQ(\xi) : \chi \in T(X)\Big\}, \\
\pi(\chi) &:= \inf\{\langle c, x\rangle : x \in X, Tx = \chi\}, \\
\sigma(\chi) &:= \text{argmin}\{\langle c, x\rangle : x \in X, Tx = \chi\} \quad (\chi \in T(X)).
\end{aligned}$$

Moreover, π is convex polyhedral on $T(X)$ and σ is a polyhedral set-valued mapping which is Lipschitz continuous on $T(X)$ with respect to the Pompeiu-Hausdorff distance.

Proof: Let $\bar{x} \in X^*(Q)$. We set $\Phi_Q(\chi) := \int_{\Xi} \Phi(q(\xi), h(\xi) - \chi)dQ(\xi)$ and have

$$\vartheta(Q) = \langle c, \bar{x}\rangle + \Phi_Q(T\bar{x}) \geq \inf\{\pi(\chi) + \Phi_Q(\chi) : \chi \in T(X)\}.$$

For the converse inequality, let $\varepsilon > 0$ and $\bar{\chi} \in T(X)$ be such that

$$\pi(\bar{\chi}) + \Phi_Q(\bar{\chi}) \leq \inf\{\pi(\chi) + \Phi_Q(\chi) : \chi \in T(X)\} + \frac{\varepsilon}{2}.$$

Then there exists an $\bar{x} \in X$ such that $T\bar{x} = \bar{\chi}$ and $\langle c, \bar{x}\rangle \leq \pi(\bar{\chi}) + \frac{\varepsilon}{2}$. Hence,

$$\begin{aligned}
\vartheta(Q) &\leq \langle c, \bar{x}\rangle + \Phi_Q(T\bar{x}) \leq \pi(\bar{\chi}) + \Phi_Q(\bar{\chi}) + \frac{\varepsilon}{2} \\
&\leq \inf\{\pi(\chi) + \Phi_Q(\chi) : \chi \in T(X)\} + \varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, the first statement is verified. In particular, $x \in \sigma(Tx)$ and $Tx \in Y^*(Q)$ for any $x \in X^*(Q)$. Hence, it holds that $X^*(Q) \subseteq \sigma(Y^*(Q))$. Conversely, let $x \in \sigma(Y^*(Q))$. Then $x \in \sigma(\chi)$ for some $\chi \in Y^*(Q)$. Thus $Tx = \chi$ and $\langle c, x\rangle = \pi(\chi) = \pi(Tx)$, implying

$$\begin{aligned}
\langle c, x\rangle + \Phi_Q(Tx) &= \pi(Tx) + \Phi_Q(Tx) = \inf\{\pi(\chi) + \Phi_Q(\chi) : \chi \in T(X)\} \\
&= \vartheta(Q) \quad \text{and} \quad x \in X^*(Q).
\end{aligned}$$

Furthermore, π is clearly convex and polyhedral, and the properties of σ are well known (cf. Walkup and Wets (1969b)). $\square$

Theorem 29 *Let (A1),(A2) be satisfied, $X^*(P)$ be nonempty and $\mathcal{U}$ be an open bounded neighbourhood of $X^*(P)$. Furthermore, assume that $T(X^*(P))$*

is a singleton. Then there exist constants $L > 0$ and $\delta > 0$ such that

$$\mathbb{D}_\infty(X^*(P), X^*(Q)) \le L\Psi_P(L\zeta_2(P,Q))$$

whenever $Q \in \mathcal{P}_2(\Xi)$ and $\zeta_2(P,Q) < \delta$, where Ψ_P is given by (23) and $\mathbb{D}_\infty$ denotes the Pompeiu-Hausdorff distance.

Proof: Let χ^* be the single element belonging to $T(X^*(P))$. We use the notation of Lemma 28 and conclude that $Y^*(P) = \{\chi^*\}$. Let $\mathcal{V}$ denote a neighbourhood of χ^* such that $T^{-1}(\mathcal{V}) \subset \mathcal{U}$ and consider the growth function

$$\psi_P^*(\tau) := \min\{\pi(\chi) + \Phi_P(\chi) - \vartheta(P) : \|\chi - \chi^*\| \ge \tau, \chi \in T(X) \cap \mathcal{V}\}$$

and the associated function $\Psi_P^*(\eta) := \eta + (\psi_P^*)^{-1}(2\eta)$ of the stochastic program $\inf\{\pi(\chi) + \Phi_P(\chi) : \chi \in T(X)\}$. Applying Corollary 14 to the latter program yields the estimate

$$\sup_{\chi \in Y^*(Q)} d(\chi, Y^*(P)) = \sup_{\chi \in Y^*(Q)} \|\chi - \chi^*\| \le \Psi_P^*(L_*\zeta_2(P,Q))$$

for some $L_* > 0$ and small $\zeta_2(P,Q)$. Since $X^*(P) = \sigma(\chi^*)$ and $X^*(Q) = \sigma(Y^*(Q))$ hold due to Lemma 28 and the set-valued mapping σ is Lipschitz continuous on $T(X)$ with respect to $\mathbb{D}_\infty$ (with some constant $L_\sigma > 0$), we obtain

$$\begin{aligned}\mathbb{D}_\infty(X^*(P), X^*(Q)) = \mathbb{D}_\infty(\sigma(\chi^*), \sigma(Y^*(Q)) &\le \sup_{\chi \in Y^*(Q)} \mathbb{D}_\infty(\sigma(\chi^*), \sigma(\chi)) \\ &\le L_\sigma \sup_{\chi \in Y^*(Q)} \|\chi^* - \chi\| \le L_\sigma \Psi_P^*(L_*\zeta_2(P,Q)).\end{aligned}$$

It remains to explore the relation between the two growth functions ψ_P and ψ_P^*, and the associated functions Ψ_P and Ψ_P^*, respectively. Let $\tau \in \mathbb{R}_+$ and $\chi_\tau \in T(X) \cap \mathcal{V}$ such that $\|\chi_\tau - \chi^*\| \ge \tau$ and $\psi_P^*(\tau) = \pi(\chi_\tau) + \Phi_P(\chi_\tau) - \vartheta(P)$. Let $x_\tau \in X$, $\tilde{x}_\tau \in X^*(P)$ be such that $Tx_\tau = \chi_\tau$, $\pi(\chi_\tau) = cx_\tau$ and $d(x_\tau, X^*) = \|x_\tau - \tilde{x}_\tau\|$. Hence, we obtain $x_\tau \in \mathcal{U}$, $\psi_P^*(\tau) = cx_\tau + \Phi_P(Tx_\tau) - \vartheta(P)$ and

$$\tau \le \|\chi_\tau - \chi^*\| = \|Tx_\tau - T\tilde{x}_\tau\| \le \|T\| d(x_\tau, X^*),$$

where $\|T\|$ denotes the matrix norm of T. If $\|T\| \neq 0$, we conclude that $\psi_P^*(\tau) \ge \psi_P(\frac{\tau}{\|T\|})$ holds for any $\tau \in \mathbb{R}_+$ and, hence, we have $(\psi_P^*)^{-1}(\eta) \le \|T\|\psi_P^{-1}(\eta)$ and $\Psi_P^*(\eta) \le \max\{1, \|T\|\}\Psi_P(\eta)$ for any $\eta \in \mathbb{R}_+$. This implies

$$\mathbb{D}_\infty(X^*(P), X^*(Q)) \le \max\{1, \|T\|\} L_\sigma \Psi_P(L_*\zeta_2(P,Q)),$$

and, thus, the desired estimate. In case of $\|T\| = 0$, the solution set $X^*(P)$ is equal to $\operatorname{argmin}\{\langle c, x\rangle : x \in X\}$ and, consequently, does not change if P is perturbed. Hence, the result is correct in the latter case, too. □

Theorem 30 *Let (A1),(A2) be satisfied, $X^*(P)$ be nonempty, $\mathcal{U}$ be an open bounded neighbourhood of $X^*(P)$ and $T(X^*(P))$ be a singleton. Assume that the function $(\psi_P^r)^{-1}$ is continuous at $\tau = 0$, where $\psi_P^r(0) = 0$, $\psi_P^r(\tau) := \frac{1}{\tau}\psi_P(\tau)$ for each $\tau > 0$ and $\psi_P(\cdot)$ is the growth function given by (22).*
Then there exists constants $L > 0$ and $\delta > 0$ such that the estimate

$$\mathbb{D}_\infty(X^*(P), X^*(Q)) \le L(\psi_P^r)^{-1}(\hat{d}_{\Phi_\mathcal{U}}(P,Q)) \tag{33}$$

is valid for each $Q \in \mathcal{P}_2(\Xi)$ with $\hat{d}_{\Phi_\mathcal{U}}(P,Q) < \delta$. Here, we denote

$$\hat{d}_{\Phi_\mathcal{U}}(P,Q) := \sup\Big\{\Big|\int_\Xi \frac{\Phi(q(\xi), h(\xi) - Tx) - \Phi(q(\xi), h(\xi) - T\bar{x})}{\|x - \bar{x}\|} d(P-Q)(\xi)\Big| : x, \bar{x} \in X \cap \operatorname{cl}\mathcal{U}, x \ne \bar{x}\Big\}.$$

If the two-stage model (31) has quadratic growth, the estimate (33) asserts Lipschitz continuity with respect to $\hat{d}_{\Phi_\mathcal{U}}$.

Proof: Using the same notation as in the previous proof we conclude again that

$$\mathbb{D}_\infty(X^*(P), X^*(Q)) \le L_\sigma \sup_{\chi \in Y^*(Q)} \|\chi^* - \chi\|.$$

If T is the null matrix, the result is true since $X^*(Q)$ does not depend on Q. Otherwise, we denote by $\|T\|$ the matrix norm of T, argue as in the proofs of the Theorems 12 and 29 and arrive at the estimate

$$\psi_P(\frac{1}{\|T\|}\|\chi - \chi^*\|) \le \psi_P^*(\|\chi - \chi^*\|) \le \Phi_P(\chi) - \Phi_Q(\chi) - (\Phi_P(\chi^*) - \Phi_Q(\chi^*))$$

for each $\chi \in Y^*(Q)$, where $\Phi_P(\chi) := \int_\Xi \Phi(q(\xi), h(\xi) - \chi) dP(\xi)$. The latter estimate implies (33). □

Remark 31 In all cases, where the original and perturbed solution sets $X^*(P)$ and $X^*(Q)$ are convex and an estimate of the form

$$\mathbb{D}_\infty(X^*(P), X^*(Q)) \le \phi(d(P,Q)) \quad \text{whenever} \quad Q \in \mathcal{P}_d,\ d(P,Q) < \delta$$

is available for some (pseudo) metric d on a set of probability measures $\mathcal{P}_d$ and some function ϕ from $\mathbb{R}_+$ to $\mathbb{R}_+$, this estimate may be complemented by a quantitative continuity property of a countable dense family of selections. Namely, there exists a family $\{x_k^*(Q)\}_{k\in\mathbb{N}}$ of selections of $X^*(Q)$ such that

$$X^*(Q) = \mathrm{cl}\Big(\bigcup_{k\in\mathbb{N}} x_k^*(Q)\Big)$$
$$\|x_k^*(P) - x_k^*(Q)\| \leq L_k \phi(d(P,Q)) \quad \text{whenever} \quad Q \in \mathcal{P}_d,\ d(P,Q) < \delta$$

for some constant $L_k > 0$ and any $k \in \mathbb{N}$. To derive this conclusion, let us first recall the notion of a generalized Steiner point of a convex compact set $C \subset \mathbb{R}^m$ (see Dentcheva (2000)). It is given by $\mathrm{St}_\alpha(C) := \int_{\mathbb{B}} \mu(\partial\sigma_C(x))\alpha(dx)$, where $\sigma_C(\cdot)$ is the support function of C, i.e., $\sigma_C(x) := \sup_{y\in C}\langle x, y\rangle$, $\partial\sigma_C(x)$ is the convex subdifferential of σ_C at x and $\mu(\partial\sigma_C(x))$ its norm-minimal element. Furthermore, α is a probability measure on $\mathbb{B}$ having a C^1-density with respect to the Lebesgue measure. A generalized Steiner selection $\mathrm{St}_\alpha(\cdot)$ is Lipschitz continuous (with a Lipschitz constant depending on α) on the set of all nonempty convex compact subsets of $\mathbb{R}^m$ equipped with the distance $\mathbb{D}_\infty$. Furthermore, there exists a countable family $\{\alpha_k\}_{k\in\mathbb{N}}$ of probability measures on $\mathbb{R}$, each having a C^1-density with respect to the Lebesgue measure, such that the corresponding family of generalized Steiner selections $\{\mathrm{St}_{\alpha_k}(C)\}_{k\in\mathbb{N}}$ is dense in C. Both results are proved in Dentcheva (2000). By combining these two arguments for the countable family $\{x_k^*(Q) := \mathrm{St}_{\alpha_k}(X^*(Q))\}_{k\in\mathbb{N}}$ of selections to the convex compact sets $X^*(Q)$ the desired result follows.

The previous Theorems 29 and 30 extend the main results of Römisch and Schultz (1993, 1996) and Shapiro (1994) to the case of a general growth condition. The crucial assumption of both results is that $T(X^*(P))$ is a singleton. The latter condition is satisfied, for example, if the expected recourse function $\Phi_P(\cdot) := \int_\Xi \Phi(q(\xi), h(\xi) - \cdot)dP(\xi)$ is strictly convex on a convex neighbourhood of $T(X^*(P))$.
The situation simplifies in case of random right-hand sides only, i.e., $q(\xi) \equiv q$ and $h(\xi) = \xi$. Then the distance $\hat{d}_{\Phi_\mathcal{U}}$ can be bounded above by a discrepancy w.r.t. certain polyhedral cones. Namely,

$$\hat{d}_{\Phi_\mathcal{U}}(P,Q) \leq \hat{L}\sup\{|(P-Q)(Tx + B_i(\mathbb{R}^s_+))| : x \in \mathrm{cl}\,\mathcal{U}, i = 1,\ldots,\ell\},$$

holds, where $\hat{L} > 0$ is some constant and B_i, $i = 1,\ldots,\ell$, are certain nonsingular submatrices of the recourse matrix W (Römisch and Schultz (1996)). In this case, verifiable sufficient conditions for the strict and strong convexity of the expected recourse function Φ_P are also available (Schultz (1994)). Namely, the function Φ_P is strictly convex on any open convex subset of the support of P if P has a density on $\mathbb{R}^s$ and the set $\{z \in \mathbb{R}^s : W'z < q\}$ is nonempty. It is strongly convex if, in addition to the conditions implying strict convexity, the density of P is bounded away from zero on the corresponding convex neighbourhood. Furthermore, the model (31) has quadratic growth if the function

Φ_P is strongly convex on some open convex neighbourhood of $T(X^*(P))$. The latter fact was proved in Dentcheva and Römisch (2000) by exploiting the Lipschitz continuity of the mapping σ in Lemma 28. The Lipschitz continuity result of Theorem 30 in case of quadratic growth forms the basis of the following differential stability result for optimal values and solution sets proved in Dentcheva and Römisch (2000).

Theorem 32 *Let (A1),(A2) be satisfied, $X^*(P)$ be nonempty and bounded, and $T(X^*(P))$ be a singleton, i.e., $T(X^*(P)) = \{\chi^*\}$. Let $Q \in \mathcal{P}(\Xi)$. Then the function ϑ is Gateaux directionally differentiable at P in direction $Q - P$ and it holds*

$$\vartheta'(P; Q-P) := \lim_{t\to 0+} \frac{1}{t}(\vartheta(P + t(Q-P)) - \vartheta(P)) = \Phi_Q(\chi^*) - \Phi_P(\chi^*).$$

If, in addition, model (31) has quadratic growth and Φ_P is twice continuously differentiable at $\{\chi^*\}$, then the second-order Gateaux directional derivative of ϑ at P in direction $Q - P$ exists and we have

$$\begin{aligned}\vartheta''(P; Q-P) &:= \lim_{t\to 0+} \frac{1}{t^2}(\vartheta(P + t(Q-P)) - \vartheta(P) - t\vartheta'(P; Q-P)) \\ &= \inf\Big\{\frac{1}{2}\langle \nabla^2\Phi_P(\chi^*)Tx, Tx\rangle + (\Phi_Q - \Phi_P)'(\chi^*; Tx) : x \in S(\bar{x})\Big\},\end{aligned}$$

where $S(\bar{x}) = \{x \in T_X(\bar{x}) : cx + \langle\nabla\Phi_P(\chi^*), Tx\rangle = 0\}$ and $T_X(\bar{x})$ is the tangent cone to X at some $\bar{x} \in X^*(P)$. The directional derivative $(\Phi_Q - \Phi_P)'(\chi^*; Tx)$ of $\Phi_Q - \Phi_P$ exists since both functions are convex and Φ_P is differentiable. The first-order Gateaux directional derivative of the set-valued mapping $X^*(\cdot)$

$$(X^*)'(P, \bar{x}; Q-P) := \lim_{t\to 0+} \frac{1}{t}(X^*(P + t(Q-P)) - \bar{x})$$

at the pair $(P, \bar{x})$, $\bar{x} \in X^*(P)$, in direction $Q - P$ exists and coincides with $\arg\min\{\frac{1}{2}\langle \nabla^2\Phi_P(\chi^*)Tx, Tx\rangle + (\Phi_Q - \Phi_P)'(\chi^*; Tx) : x \in S(\bar{x})\}$.

3.2 Mixed-Integer Two-Stage Models

Next we allow for mixed-integer decisions in both stages and consider the stochastic program

$$\min\{\langle c, x\rangle + \int_\Xi \Phi(h(\xi) - T(\xi)x)dP(\xi) : x \in X\}, \tag{34}$$

where

$$\Phi(t) := \min\{\langle q, y\rangle + \langle \bar{q}, \bar{y}\rangle : Wy + \bar{W}\bar{y} = t, y \in \mathbb{Z}_+^{\hat{m}}, \bar{y} \in \mathbb{R}_+^{\bar{m}}\}\ (t \in \mathbb{R}^r), \quad (35)$$

$c \in \mathbb{R}^m$, X is a closed subset of $\mathbb{R}^m$, Ξ a polyhedron in $\mathbb{R}^s$, $q \in \mathbb{R}^{\hat{m}}$, $\bar{q} \in \mathbb{R}^{\bar{m}}$, W and $\bar{W}$ are $(r, \hat{m})$- and $(r, \bar{m})$-matrices, respectively, $h(\xi) \in \mathbb{R}^r$ and the (r, m)-matrix $T(\xi)$ are affine linear functions of $\xi \in \mathbb{R}^s$, and $P \in \mathcal{P}(\Xi)$.
Basic properties of Φ like convexity and continuity on $\mathrm{dom}\,\Phi$ in the purely linear case cannot be maintained for reasonable problem classes. Since Φ is discontinuous in general it is interesting to characterize its continuity regions. Similarly as for the two-stage models without integrality requirements in the previous section, we need some conditions to have the model (34) well-defined:
(B1) The matrices W and $\bar{W}$ have only rational elements.
(B2) For each pair $(x, \xi) \in X \times \Xi$ it holds that $h(\xi) - T(\xi)x \in \mathcal{T}$, where $\mathcal{T} := \{t \in \mathbb{R}^r : t = Wy + \bar{W}\bar{y},\ y \in \mathbb{Z}_+^{\hat{m}},\ \bar{y} \in \mathbb{R}_+^{\bar{m}}\}$.
(B3) There exists an element $u \in \mathbb{R}^r$ such that $W'u \leq q$ and $\bar{W}'u \leq \bar{q}$.
(B4) $P \in \mathcal{P}_1(\Xi)$, i.e., $\int_\Xi \|\xi\| dP(\xi) < +\infty$.
The conditions (B2) and (B3) mean *relatively complete recourse* and *dual feasibility*, respectively. We note that condition (B3) is equivalent to $\Phi(0) = 0$, and that (B2) and (B3) imply $\Phi(t)$ to be finite for all $t \in \mathcal{T}$ (see Proposition 1 in Louveaux and Schultz (2003)). In the context of this section, the following properties of the value function Φ on $\mathcal{T}$ are important.

Lemma 33 *Assume (B1)–(B3). Then there exists a countable partition of $\mathcal{T}$ into Borel subsets $\mathcal{B}_i$, i.e., $\mathcal{T} = \bigcup_{i\in\mathbb{N}} \mathcal{B}_i$ such that*

(1) each of the sets has a representation $\mathcal{B}_i = \{b_i + \mathrm{pos}\,\bar{W}\} \setminus \bigcup_{j=1}^{N_0} \{b_{ij} + \mathrm{pos}\,\bar{W}\}$, where $b_i, b_{ij} \in \mathbb{R}^r$ for $i \in \mathbb{N}$ and $j = 1, \ldots, N_0$. Moreover, there exists an $N_1 \in \mathbb{N}$ such that for any $t \in \mathcal{T}$ the ball $\mathbb{B}(t, 1)$ in $\mathbb{R}^r$ is intersected by at most N_1 different subsets $\mathcal{B}_i$.

(2) the restriction $\Phi|_{\mathcal{B}_i}$ of Φ to $\mathcal{B}_i$ is Lipschitz continuous with a constant $L_\Phi > 0$ that does not depend on i.

Furthermore, the function Φ is lower semicontinuous and piecewise polyhedral on $\mathcal{T}$ and there exist constants $a, b > 0$ such that it holds for all $t, \tilde{t} \in \mathcal{T}$:

$$|\Phi(t) - \Phi(\tilde{t})| \leq a\|t - \tilde{t}\| + b\,.$$

Part (i) of the lemma was proved in Section 5.6 of Bank et al. (1982) and in Lemma 2.5 of Schultz (1996), (ii) was derived as Lemma 2.3 in Schultz (1996) and the remaining properties of Φ were established in Blair and Jeroslow (1977). Compared to Lemma 21 for optimal value functions of linear programs without integrality requirements, the representation of Φ is now given on countably many (possibly unbounded) Borel sets. This requires to incorporate the tail behaviour of P and leads to the following representation of the function $F_0(x,\xi) := \langle c,x\rangle + \Phi(h(\xi) - T(\xi)x)$ for each pair (x,ξ) in $X \times \Xi$.

Proposition 34 *Assume (B1)–(B3) and let $\mathcal{U}$ be an open bounded subset of $\mathbb{R}^m$. For each $R \geq 1$ and $x \in X \cap \mathrm{cl}\,\mathcal{U}$ there exist disjoint Borel subsets $\Xi^R_{j,x}$ of Ξ, $j = 1,\ldots,\nu$, whose closures are polyhedra with a uniformly bounded number of faces such that the function*

$$F_0(x,\xi) = \sum_{j=0}^{\nu} (\langle c,x\rangle + \Phi(h(\xi) - T(\xi)x))\chi_{\Xi^R_{j,x}}(\xi) \quad ((x,\xi) \in X \times \Xi)$$

is Lipschitz continuous with respect to ξ on each $\Xi^R_{j,x}$, $j = 1,\ldots,\nu$, with some uniform Lipschitz constant. Here, $\Xi^R_{0,x} := \Xi \setminus \cup_{j=1}^{\nu}\Xi^R_{j,x}$ is contained in $\{\xi \in \mathbb{R}^s : \|\xi\| > R\}$ and ν is bounded by a multiple of R^r.

Proof: Since $h(\cdot)$ and $T(\cdot)$ are affine linear functions, there exists a constant $C_2 > 0$ such that the estimate $\|h(\xi) - T(\xi)x\|_\infty \leq C_2 \max\{1, \|\xi\|\}$ holds for each pair in $X \cap \mathrm{cl}\,\mathcal{U}$. Let $R > 0$ and $\mathcal{T}_R := \mathcal{T} \cap RC_2\mathbb{B}_\infty$, where $\mathbb{B}_\infty$ refers to the closed unit ball in $\mathbb{R}^r$ with respect to the norm $\|\cdot\|_\infty$. Now, we partition the ball $RC_2\mathbb{B}_\infty$ into disjoint Borel sets whose closures are $\mathbb{B}_\infty$-balls with radius 1, where possible gaps are filled with maximal balls of radius less than 1. Then the number of elements in this partition of $RC_2\mathbb{B}_\infty$ is bounded above by $(2RC_2)^r$. From Lemma 33 (i) we know that each element of this partition is intersected by at most N_1 subsets $\mathcal{B}_i$ (for some $N_1 \in \mathbb{N}$). Another consequence of Lemma 33 (i) is that each $\mathcal{B}_i$ splits into disjoint Borel subsets whose closures are polyhedra. Moreover, the number of such subsets can be bounded from above by a constant not depending on i. Hence, there exist a number $\nu \in \mathbb{N}$ and disjoint Borel subsets $\{B_j : j = 1,\ldots,\nu\}$ such that their closures are polyhedra, their union contains $\mathcal{T}_R$, and ν is bounded above by κR^r, where the constant $\kappa > 0$ is independent of R. Now, let $x \in X \cap \mathrm{cl}\,\mathcal{U}$ and consider the following disjoint Borel subsets of Ξ:

$$\Xi^R_{j,x} := \{\xi \in \Xi : h(\xi) - T(\xi)x \in B_j\} \quad (j = 1,\ldots,\nu),$$

$$\Xi^R_{0,x} := \Xi \setminus \bigcup_{j=1}^{\nu} \Xi^R_{j,x} \subseteq \{\xi \in \Xi : \|h(\xi) - T(\xi)x\|_\infty > RC_2\} \subseteq \{\xi \in \Xi : \|\xi\| > R\}.$$

For each $j = 1,\ldots,\nu$ the closures of the sets B_j are polyhedra with a number of faces that is bounded above by some number not depending on j, ν and R. Hence, the same is true for the closures of the sets $\Xi^R_{j,x}$, i.e., for $\{\xi \in \Xi : h(\xi) - T(\xi)x \in \mathrm{cl}\,B_j\}$, where, moreover, the corresponding number $k \in \mathbb{N}$ does not depend on $x \in X \cap \mathrm{cl}\,\mathcal{U}$. Finally, we conclude from Lemma 33 that there exists a constant $L_1 > 0$ (which does not depend on $x \in X \cap \mathrm{cl}\,\mathcal{U}$, $j = 1,\ldots,\nu$ and $R > 0$) such that the function $F_0(x,\cdot)|_{\Xi^R_{j,x}} = \langle c,x\rangle + \Phi|_{B_j}(h(\cdot) - T(\cdot)x)$ is Lipschitz continuous with constant L_1. $\square$

For further structural properties of model (34) we refer to Louveaux and Schultz (2003). In order to state stability results for model (34), we consider the following probability metrics with ζ-structure on $\mathcal{P}_1(\Xi)$ for every $k \in \mathbb{N}$:

$$\begin{aligned}\zeta_{1,\mathrm{ph}_k}(P,Q) &:= \sup\{|\int_B f(\xi)(P-Q)(d\xi)| : f \in \mathcal{F}_1(\Xi), B \in \mathcal{B}_{\mathrm{ph}_k}(\Xi)\} \qquad (36)\\ &= \sup\{|\int_\Xi f(\xi)\chi_B(\xi)(P-Q)(d\xi)| : f \in \mathcal{F}_1(\Xi), B \in \mathcal{B}_{\mathrm{ph}_k}(\Xi)\}.\end{aligned}$$

Here, $\mathcal{B}_{\mathrm{ph}_k}(\Xi)$ and $\mathcal{F}_1(\Xi)$ denote the sets of polyhedra in Ξ and of Lipschitz continuous functions from Ξ to $\mathbb{R}$ introduced in Section 2.1.

Theorem 35 *Let the conditions (B1)–(B4) be satisfied, $X^*(P)$ be nonempty and $\mathcal{U} \subseteq \mathbb{R}^m$ be an open bounded neighbourhood of $X^*(P)$.*
Then there exist constants $L > 0$, $\delta > 0$ and $k \in \mathbb{N}$ such that

$$\begin{aligned}|\vartheta(P) - \vartheta_{\mathcal{U}}(Q)| &\le L\phi_P(\zeta_{1,\mathrm{ph}_k}(P,Q)) \qquad (37)\\ \emptyset \neq X^*_{\mathcal{U}}(Q) &\subseteq X^*(P) + \Psi_P(L\phi_P(\zeta_{1,\mathrm{ph}_k}(P,Q)))\mathbb{B},\end{aligned}$$

and $X^*_{\mathcal{U}}(Q)$ is a CLM set of (34) relative to $\mathcal{U}$ whenever $Q \in \mathcal{P}_1(\Xi)$ and $\zeta_{1,\mathrm{ph}_k}(P,Q) < \delta$. Here, the function ϕ_P on $\mathbb{R}_+$ is defined by

$$\phi_P(0) = 0 \quad \text{and} \quad \phi_P(t) := \inf_{R \ge 1}\Big\{R^r t + \int_{\{\xi\in\Xi:\|\xi\|>R\}} \|\xi\| dP(\xi)\Big\} \quad (t > 0)$$

and continuous at $t = 0$, and the function Ψ_P is given by (23).
If P has a finite absolute moment of p-th order for some $p > 1$, the estimate $\phi_P(t) \le Ct^{\frac{p-1}{p-1+r}}$ holds for small $t > 0$ and some constant $C > 0$.

Proof: Since the function Φ is lower semicontinuous on $\mathcal{T}$ (Lemma 33), F_0 is lower semicontinuous on $X \times \Xi$ and, hence, a random lower semicontinuous function (Example 14.31 in Rockafellar and Wets (1998)). Using Lemma 33 we obtain the estimate

$$|F_0(x,\xi)| \le \|c\|\|x\| + a(\|h(\xi)\| + \|T(\xi)\|\|x\|) + b$$

for each pair $(x,\xi) \in X \times \Xi$. Since $h(\xi)$ and $T(\xi)$ depend affine linearly on ξ, there exists a constant $C_1 > 0$ such that $|F_0(x,\xi)| \le C_1 \max\{1, \|\xi\|\}$ holds for each pair $(x,\xi) \in (X \cap \mathrm{cl}\mathcal{U}) \times \Xi$. Hence, $\mathcal{P}_{\mathcal{F}_{\mathcal{U}}}(\Xi) \supseteq \mathcal{P}_1(\Xi)$ and Theorems 5 and 9 apply with $d = 0$ and the distance $d_{\mathcal{F}_{\mathcal{U}}}$ on $\mathcal{P}_1(\Xi)$.
From Proposition 34 we know that, for each $R \ge 1$ and $x \in X \cap \mathrm{cl}\mathcal{U}$, there exist Borel subsets $\Xi^R_{j,x}$, $j = 1, \ldots, \nu$, of Ξ such that the function $f^R_{j,x}(\cdot) := F_0(x,\cdot)|_{\Xi^R_{j,x}}$ is Lipschitz continuous on $\Xi^R_{j,x}$ with some Lipschitz constant $L_1 > 0$

(not depending on x, j and R). We extend each function $f_{j,x}^R(\cdot)$ to the whole of Ξ by preserving the Lipschitz constant L_1. Proposition 34 also implies that the closures of $\Xi_{j,x}^R$ are contained in $\mathcal{B}_{\mathrm{ph}_k}(\Xi)$ for some $k \in \mathbb{N}$, that the number ν is bounded above by κR^r, where the constant $\kappa > 0$ is independent on R, and that $\Xi_{0,x}^R := \Xi \setminus \cup_{j=1}^{\nu} \Xi_{j,x}^R$ is a subset of $\{\xi \in \Xi : \|\xi\| > R\}$.
For each $Q \in \mathcal{P}_1(\Xi)$ and $x \in X \cap \mathrm{cl}\,\mathcal{U}$ we obtain

$$
\begin{aligned}
\Big|\int_{\Xi} F_0(x,\xi)d(P-Q)(\xi)\Big| &= \Big|\sum_{j=0}^{\nu} \int_{\Xi_{j,x}^R} F_0(x,\xi)d(P-Q)(\xi)\Big| \\
&\le \sum_{j=1}^{\nu} \Big|\int_{\Xi_{j,x}^R} f_{j,x}^R(\xi)d(P-Q)(\xi)\Big| + I_x^R(P,Q) \\
&\le \nu L_1 \sup_{\substack{f\in\mathcal{F}_1(\Xi)\\ j=1,\dots,\nu}} \Big|\int_{\Xi} f(\xi)\chi_{\Xi_{j,x}^R} d(P-Q)(\xi)\Big| + I_x^R(P,Q),
\end{aligned}
$$

where $I_x^R(P,Q) := |\int_{\Xi_{0,x}^R} F_0(x,\xi)d(P-Q)(\xi)|$.
For each $\Xi_{j,x}^R$ we now consider a sequence of polyhedra $B_{j,x}^R$, which are contained in $\Xi_{j,x}^R$ and belong to $\mathcal{B}_{\mathrm{ph}_k}(\Xi)$, such that their characteristic functions $\chi_{B_{j,x}^R}$ converge pointwise to the characteristic function $\chi_{\Xi_{j,x}^R}$. Then the sequence consisting of the elements $|\int_\Xi f(\xi)\chi_{B_{j,x}^R}(\xi)d(P-Q)(\xi)|$ converges to $|\int_\Xi f(\xi)\chi_{\Xi_{j,x}^R}(\xi)d(P-Q)(\xi)|$ while each element is bounded by $\zeta_{1,\mathrm{ph}_k}(P,Q)$. Hence, the above estimate may be continued to

$$
\Big|\int_{\Xi} F_0(x,\xi)d(P-Q)(\xi)\Big| \le \kappa L_1 R^r \zeta_{1,\mathrm{ph}_k}(P,Q) + I_x^R(P,Q). \tag{38}
$$

For the term $I_x^R(P,Q)$ we have

$$
\begin{aligned}
I_x^R(P,Q) &\le C_1 \int_{\{\xi\in\Xi:\|\xi\|>R\}} \|\xi\| d(P+Q)(\xi) \\
&\le C_1 \int_{\{\xi\in\Xi:\|\xi\|_\infty > \frac{R}{C_2}\}} \|\xi\| d(P+Q)(\xi)
\end{aligned}
$$

where we have used the estimate $|F_0(x,\xi)| \le C_1\|\xi\|$ for each pair $(x,\xi) \in (X \cap \mathrm{cl}\,\mathcal{U}) \times \{\xi \in \Xi : \|\xi\| > R\}$ and $C_2 > 0$ is a norming constant such that $\|\xi\| \le C_2\|\xi\|_\infty$ holds for each $\xi \in \mathbb{R}^s$. Clearly, the set $\{\xi \in \Xi : \|\xi\|_\infty > \frac{R}{C_2}\}$ can be covered by 2^s intersections of Ξ by open halfspaces whose closures belong to $\mathcal{B}_{\mathrm{ph}_k}(\Xi)$. Hence, a similar argument as the one above yields the estimate

$$\int_{\{\xi\in\Xi:\|\xi\|_\infty>\frac{R}{C_2}\}} \|\xi\| dQ(\xi) \le 2^s \zeta_{1,\mathrm{ph}_k}(P,Q) + \int_{\{\xi\in\Xi:\|\xi\|_\infty>\frac{R}{C_2}\}} \|\xi\| dP(\xi).$$

Hence, from the previous estimates we obtain that

$$\begin{aligned} d_{\mathcal{F}_\mathcal{U}}(P,Q) &\le \kappa(L_1 R^r + 2^s C_1)\zeta_{1,\mathrm{ph}_k}(P,Q) + 2C_1 \int_{\{\xi\in\Xi:\|\xi\|_\infty>\frac{R}{C_2}\}} \|\xi\| dP(\xi) \\ &\le C R^r \zeta_{1,\mathrm{ph}_k}(P,Q) + \int_{\{\xi\in\Xi:\|\xi\|>\alpha R\}} \|\xi\| dP(\xi) \end{aligned}$$

for some constants $C > 0$ and $\alpha \in (0,1)$, the latter depending on the norming constants of $\|\cdot\|$ and $\|\cdot\|_\infty$, respectively. Finally, we obtain

$$d_{\mathcal{F}_\mathcal{U}}(P,Q) \le \hat{C}\phi_P(\zeta_{1,\mathrm{ph}_k}(P,Q)), \quad \text{where} \tag{39}$$

$$\phi_P(0) := 0 \quad \text{and} \quad \phi_P(t) := \inf_{R\ge 1} \Big\{ R^r t + \int_{\{\xi\in\Xi:\|\xi\|>R\}} \|\xi\| dP(\xi) \Big\} \; (t > 0) \tag{40}$$

with some constant $\hat{C} > 0$. Now, the result is a consequence of the Theorems 5 and 9. If $\int_\Xi \|\xi\|^p dP(\xi) < \infty$, it holds that $\int_{\{\xi\in\Xi:\|\xi\|>R\}} \|\xi\| dP(\xi) \le R^{1-p}\int_\Xi \|\xi\|^p dP(\xi)$ by Markov's inequality. The desired estimate follows by inserting $R = t^{-\frac{1}{p+r-1}}$ for small $t > 0$ into the function whose infimum w.r.t. $R \ge 1$ is $\phi_P(t)$. $\square$

In case that the underlying distribution P and its perturbations Q have supports in some bounded subset of $\mathbb{R}^s$, the stability result improves slightly.

Corollary 36 *Let the conditions (B1)–(B3) be satisfied and Ξ be bounded. Assume that $P \in \mathcal{P}(\Xi)$, $X^*(P)$ is nonempty and $\mathcal{U} \subseteq \mathbb{R}^m$ is an open bounded neighbourhood of $X^*(P)$.*
Then there exist constants $L > 0$, $\delta > 0$ and $k \in \mathbb{N}$ such that

$$\begin{aligned} |\vartheta(P) - \vartheta_\mathcal{U}(Q)| &\le L\zeta_{1,\mathrm{ph}_k}(P,Q) \\ \emptyset \ne X^*_\mathcal{U}(Q) &\subseteq X^*(P) + \Psi_P(L\zeta_{1,\mathrm{ph}_k}(P,Q))\mathbb{B}, \end{aligned}$$

and $X^*_\mathcal{U}(Q)$ is a CLM set of (34) relative to $\mathcal{U}$ whenever $Q \in \mathcal{P}(\Xi)$ and $\zeta_{1,\mathrm{ph}_k}(P,Q) < \delta$.

Proof: Since Ξ is bounded, we have $\mathcal{P}_1(\Xi) = \mathcal{P}(\Xi)$. Moreover, the function $\phi_P(t)$ can be estimated by $R^r t$ for some sufficiently large $R > 0$. Hence, Theorem 35 implies the assertion. □

Remark 37 Since $\Xi \in \mathcal{B}_{\text{ph}_k}(\Xi)$ for some $k \in \mathbb{N}$, we obtain from (36) by choosing $B := \Xi$ and $f \equiv 1$, respectively,

$$\max\{\zeta_1(P,Q), \alpha_{\text{ph}_k}(P,Q)\} \le \zeta_{1,\text{ph}_k}(P,Q) \tag{41}$$

for large k and all $P, Q \in \mathcal{P}_1(\Xi)$. Here, α_{ph_k} denotes the polyhedral discrepancy (see Section 2.1). Hence, convergence with respect to ζ_{1,ph_k} implies weak convergence, convergence of first order absolute moments and convergence with respect to the polyhedral discrepancy α_{ph_k}. The converse is also true. The latter observation is a consequence of the estimate

$$\zeta_{1,\text{ph}_k}(P,Q) \le C_s \alpha_{\text{ph}_k}(P,Q)^{\frac{1}{s+1}} \quad (P, Q \in \mathcal{P}(\Xi)) \tag{42}$$

for some constant $C_s > 0$. It is valid for bounded $\Xi \subset \mathbb{R}^s$ and can be derived by using the technique in the proof of Proposition 3.1 in Schultz (1996). In view of (41), (42) the metric ζ_{1,ph_k} is stronger than α_{ph_k} in general, but in case of bounded Ξ both metrize the same topology on $\mathcal{P}(\Xi)$.

For more specific models (34), improvements of the above results are possible. The potential of such improvements consists in exploiting specific recourse structures, i.e., in additional information on the shape of the sets $\mathcal{B}_i$ in Lemma 33 and on the behaviour of the (value) function Φ on these sets. These considerations may lead to stability results with respect to probability metrics that are (much) weaker than ζ_{1,ph_k}. To illustrate such an improvement, let us consider the case of pure integer recourse where Φ is given by

$$\Phi(t) = \min\{\langle q, y\rangle : Wy \ge t, y \in \mathbb{Z}^{\hat{m}}\}, \tag{43}$$

the technology matrix is fixed and the right-hand side is fully stochastic, i.e., $T(\xi) \equiv T$ and $h(\xi) \equiv \xi$. This situation fits into the general model (34) by setting $\bar{q} = 0$, $\bar{m} = r$ and $\bar{W} = -I_r$, with I_r denoting the (r,r)-identity matrix. For such models Schultz (1996) observed that stability holds with respect to the Kolmogorov metric d_{K} on $\mathcal{P}(\Xi)$.

Corollary 38 *Let Φ be given by (43), $T(\xi) \equiv T$, $h(\xi) \equiv \xi$ and Ξ be bounded. Furthermore, let the conditions (B1)–(B3) be satisfied with $\mathcal{T} = \mathbb{R}^s$. Assume that $P \in \mathcal{P}(\Xi)$, $X^*(P)$ is nonempty and $\mathcal{U} \subseteq \mathbb{R}^m$ is an open bounded neighbourhood of $X^*(P)$. Then there exist constants $L > 0$ and $\delta > 0$ such that*

$$|\vartheta(P) - \vartheta_{\mathcal{U}}(Q)| \le L d_{\text{K}}(P,Q)$$
$$\emptyset \ne X^*_{\mathcal{U}}(Q) \subseteq X^*(P) + \Psi_P(L d_{\text{K}}(P,Q))\mathbb{B},$$

and $X^*_{\mathcal{U}}(Q)$ is a CLM set of (34) relative to $\mathcal{U}$ whenever $Q \in \mathcal{P}(\Xi)$ and $d_{\mathrm{K}}(P,Q) < \delta$. Here, the function Ψ_P is given by (23).

Proof: The assumptions imply that Φ is even constant on $\mathcal{B}_i$ for each $i \in \mathbb{N}$ and the continuity regions of Φ are rectangular (see Schultz (1996)). Without loss of generality the set Ξ may be chosen to be rectangular. Then the sets $\Xi^R_{j,x}$ in Proposition 34 are also bounded rectangular sets and $F_0(x,\cdot)$ is constant on each $\Xi^R_{j,x}$. Hence, the estimate (38) takes the form

$$\Big|\int_{\Xi} F_0(x,\xi)d(P-Q)(\xi)\Big| \le \kappa L_1 R^s \alpha_{\mathrm{box}}(P,Q),$$

where $\alpha_{\mathrm{box}}(P,Q) := \sup\{|P(B)-Q(B)| : B \text{ is a box in } \mathbb{R}^s\}$. Finally, we use the known estimate

$$\alpha_{\mathrm{box}}(P,Q) \le C d_{\mathrm{K}}(P,Q)$$

for some constant $C > 0$ and derive the result from Theorem 35. □

3.3 Linear Chance Constrained Programs

In this section, we study consequences of the general stability analysis of Section 2 for linear chance constrained stochastic programs of the form

$$\min\{\langle c,x\rangle : x \in X, P(\{\xi \in \Xi : T(\xi)x \ge h(\xi)\}) \ge p\}, \tag{44}$$

where $c \in \mathbb{R}^m$, X and Ξ are polyhedra in $\mathbb{R}^m$ and $\mathbb{R}^s$, respectively, $p \in (0,1)$, $P \in \mathcal{P}(\Xi)$, and the right-hand side $h(\xi) \in \mathbb{R}^r$ and the (r,m)-matrix $T(\xi)$ depend affine linearly on $\xi \in \Xi$.
We set $d = 1$, $F_0(x,\xi) = \langle c,x\rangle$, $F_1(x,\xi) = p - \chi_{H(x)}(\xi)$, where $H(x) = \{\xi \in \Xi : T(\xi)x \ge h(\xi)\}$ and $\chi_{H(x)}$ its characteristic function, and observe that the program (44) is a special case of the general stochastic program (1). We note that the set $H(x)$ is polyhedral for each $x \in X$. In fact, these sets are given as the finite intersection of r closed half-spaces. Furthermore, the multifunction H from $\mathbb{R}^m$ to $\mathbb{R}^s$ has a closed graph and, hence, the mapping $(x,\xi) \mapsto \chi_{H(x)}(\xi)$ from $\mathbb{R}^m \times \Xi$ to $\mathbb{R}$ is upper semicontinuous. This implies that F_1 is lower semicontinuous on $\mathbb{R}^m \times \Xi$ and, hence, a random lower semicontinuous function (Example 14.31 in Rockafellar and Wets (1998)). Moreover, we have $p-1 \le F_1(x,\xi) \le p$ for any pair (x,ξ). By specifying the general class of probability measures and the minimal information probability metric in Section 2.2 we obtain

$$\mathcal{P}_{\mathcal{F}_{\mathcal{U}}}(\Xi) = \Big\{Q \in \mathcal{P}(\Xi) : \sup_{x \in X \cap \mathrm{cl}\,\mathcal{U}} \max_{j=0,1} \Big|\int_{\Xi} F_j(x,\xi)dQ(\xi)\Big| < \infty\Big\} = \mathcal{P}(\Xi)$$

$$d_{\mathcal{F}_\mathcal{U}}(P,Q) = \sup_{x\in X\cap \mathrm{cl}\,\mathcal{U}} \max_{j=0,1} \Big| \int_\Xi F_j(x,\xi)(P-Q)(d\xi) \Big|$$
$$= \sup_{x\in X\cap \mathrm{cl}\,\mathcal{U}} |P(H(x)) - Q(H(x))|$$

for each $P, Q \in \mathcal{P}(\Xi)$ and any nonempty, open and bounded subset $\mathcal{U}$ of $\mathbb{R}^m$. Due to the polyhedrality of the sets $H(x)$ for any $x \in \mathbb{R}^m$, the polyhedral discrepancies α_{ph_k} on $\mathcal{P}(\Xi)$ for every $k \in \mathbb{N}$ (see Section 2.1) or related discrepancies appear as natural candidates for suitable probability metrics in case of model (44). The following result is an immediate consequence of the general methodology in Section 2.

Theorem 39 *Let $P \in \mathcal{P}(\Xi)$ and assume that*

(i) $X^*(P) \neq \emptyset$ and $\mathcal{U} \subseteq \mathbb{R}^m$ is an open bounded neighbourhood of $X^*(P)$,
(ii) the mapping $x \mapsto \{y \in \mathbb{R} : P(\{\xi \in \Xi : T(\xi)x \geq h(\xi)\}) \geq p - y\}$ is metrically regular at each pair $(\bar{x}, 0)$ with $\bar{x} \in X^*(P)$.

Then there exist constants $L > 0$, $\delta > 0$ and $k \in \mathbb{N}$ such that

$$|\vartheta(P) - \vartheta_\mathcal{U}(Q)| \leq L\,\alpha_{\mathrm{ph}_k}(P,Q)$$
$$\emptyset \neq X^*_\mathcal{U}(Q) \subseteq X^*(P) + \Psi_P(L\alpha_{\mathrm{ph}_k}(P,Q))\mathbb{B},$$

and $X^*_\mathcal{U}(Q)$ is a CLM set of (44) relative to $\mathcal{U}$ whenever $Q \in \mathcal{P}(\Xi)$ and $\alpha_{\mathrm{ph}_k}(P,Q) < \delta$. Here, the function Ψ_P is given by (23).

Proof: All sets $H(x)$ are polyhedra in $\mathbb{R}^s$ given by r linear inequalities. Hence, the number of faces of $H(x)$ is bounded by some $k \in \mathbb{N}$ not depending on $x \in \mathbb{R}^m$. Since all assumptions of Theorem 5 are satisfied for the special situation considered here, the result follows from the Theorems 5 and 9 by taking into account the estimate $d_{\mathcal{F}_\mathcal{U}}(P,Q) \leq \alpha_{\mathrm{ph}_k}(P,Q)$. □

We show that Theorem 39 applies to many chance constrained models known from the literature. First we discuss the metric regularity property (ii) of the original probabilistic constraint in (44). The following example shows that condition (ii) is indispensable for Theorem 39 to hold.

Example 40 Let $P \in \mathcal{P}(\mathbb{R})$ have a distribution function F_P which is continuously differentiable and satisfies $F_P(x) = x^{2s+1} + p$ for all x in a neighbourhood of $x = 0$ and some $p \in (0,1)$ and $s \in \mathbb{N}$. Let us consider the model

$$\min\{x : x \in \mathbb{R}, P(\xi \leq x) = F_P(x) \geq p\}.$$

Then the condition $\nabla F_P(\bar{x}) \neq 0$ is necessary and sufficient for the metric regularity at $\bar{x}$ with $F_P(\bar{x}) = p$ (Example 9.44 in Rockafellar and Wets (1998)).

Clearly, this condition is violated at the minimizer $\bar{x} = 0$. To show that the result gets lost, we consider the measures $P_n = (1 - \frac{1}{n})P + \frac{1}{n}\delta_{\frac{1}{n}}$, $n \in \mathbb{N}$. The sequence (P_n) converges weakly to P and, thus, it converges with respect to the Kolmogorov metric d_{K} as P is continuous. Then $|\vartheta(P) - \vartheta(P_n)| = (\frac{p}{n-1})^{\frac{1}{2s+1}} =: x_n$, but $d_{\mathrm{K}}(P, P_n) \geq |F_P(x_n) - F_{P_n}(x_n)| = \frac{p}{n-1}$.

When looking for general conditions implying (ii), one has to resort to results for nonconvex and nondifferentiable situations. The function

$$g(x) := P(\{\xi \in \Xi : T(\xi)x \geq h(\xi)\})$$

from $\mathbb{R}^m$ into $\mathbb{R}$ is known to be upper semicontinuous (Proposition 3.1 in Römisch and Schultz (1991c)). However, g happens to be nondifferentiable or even discontinuous not only in cases where the probability distribution P is discrete, but even if $T(\xi)$ is non-stochastic and P is continuous.

Example 41 Let P be the standard normal distribution with distribution function Φ. First let $T(\xi) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $h(\xi) = \begin{pmatrix} \xi \\ 0 \end{pmatrix}$ for each $\xi \in \mathbb{R}$. Then

$$g(x) = P(\{\xi \in \mathbb{R} : x \geq \xi, x \geq 0\}) = \begin{cases} 0, \; x < 0 \\ \Phi(x), \; x \geq 0 \end{cases}.$$

Secondly, let $T(\xi) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $h(\xi) = \begin{pmatrix} \xi \\ \xi \end{pmatrix}$ for each $\xi \in \mathbb{R}$. Then we have

$$g(x) = P(\{\xi \in \mathbb{R} : x \geq \xi, -x \geq \xi\}) = \Phi(\min\{-x, x\}).$$

We also refer to Example 9 in Henrion and Römisch (1999) for a probability distribution P having a (bounded) continuous density on $\Xi = \mathbb{R}^2$, but a probability distribution function (i.e., g in case of $T(\xi) = I$ and $h(\xi) = \xi$) that is not locally Lipschitz continuous.
Hence, one has to go back to tools from nonsmooth analysis in general. For example, if the function g is locally Lipschitz continuous on $\mathbb{R}^m$, condition (ii) is satisfied if the constraint qualification

$$\partial(-g)(\bar{x}) \cap (-N_X(\bar{x})) = \emptyset \tag{45}$$

holds at each $\bar{x} \in X^*(P)$ with $g(\bar{x}) = p$ (Corollary 4.2 in Mordukhovich (1994b)). Here, the symbol ∂ stands for the Mordukhovich subdifferential (cf.

Mordukhovich (1994a)) and $N_X(\bar{x}) := \{x^* \in \mathbb{R}^m : \langle x^*, x - \bar{x}\rangle \le 0, \forall x \in X\}$ is the normal cone to the polyhedral set X at $\bar{x} \in X$.
For more specific structures of probabilistic constraints, even in case of a stochastic matrix $T(\xi)$, the situation may become much more comfortable if P is a multivariate normal distribution. To demonstrate this, we consider the case $\Xi = \mathbb{R}^{m+1}$, $T(\xi)x = \sum_{i=1}^{m} \xi_i x_i$, i.e., $T(\xi)$ consists of one single row, and $h(\xi) = \xi_{m+1}$. Then $H(x)$ takes the form

$$H(x) = \Big\{\xi \in \mathbb{R}^{m+1} : \sum_{i=1}^{m} \xi_i x_i \ge \xi_{m+1}\Big\} \tag{46}$$

for each $x \in \mathbb{R}^m$, i.e., the sets $H(x)$ are closed half-spaces in $\mathbb{R}^{m+1}$.

Corollary 42 *Let P be a normal distribution on $\mathbb{R}^{m+1}$ with mean $\mu \in \mathbb{R}^{m+1}$ and nonsingular covariance matrix $\Sigma \in \mathbb{R}^{(m+1)\times(m+1)}$, H be given by (46) and $p \in (\frac{1}{2}, 1)$. Let $X^*(P)$ be nonempty and $\mathcal{U} \subseteq \mathbb{R}^m$ be an open bounded neighbourhood of $X^*(P)$. Assume that there exists an $\hat{x} \in X$ such that $P(H(\hat{x})) > p$. Then there are constants $L > 0$ and $\delta > 0$ such that*

$$\begin{aligned} &|\vartheta(P) - \vartheta_{\mathcal{U}}(Q)| \le L\,\alpha_h(P, Q) \\ &\qquad \emptyset \ne X^*_{\mathcal{U}}(Q) \subseteq X^*(P) + \Psi_P(L\alpha_h(P, Q))\mathbb{B} \end{aligned}$$

holds and $X^*_{\mathcal{U}}(Q)$ is a CLM set for (44) relative to $\mathcal{U}$ for each $Q \in \mathcal{P}(\Xi)$ with $\alpha_h(P, Q) < \delta$. Here, the function Ψ_P is given by (23) and α_h is the half-space discrepancy (see Section 2.1).

Proof: For any $x \in \mathbb{R}^m$, we set $x' := (x_1, \dots, x_m, -1)$ and $\sigma(x) := \langle \Sigma x', x'\rangle^{\frac{1}{2}}$. Let Φ denote the standard normal distribution function and ϕ the standard normal density. Then $\langle \xi, x'\rangle$ is normal with mean $\langle \mu, x'\rangle$ and standard deviation $\sigma(x') > 0$ (due to the nonsingularity of Σ), and

$$g(x) = P(\{\xi \in \mathbb{R}^{m+1} : \langle \xi, x'\rangle \ge 0\}) = \Phi\left(\frac{\langle \mu, x'\rangle}{\sigma(x')}\right)$$

holds for any $x \in \mathbb{R}^m$. Further, the function

$$\hat{g}(x) := \langle \mu, x'\rangle - \Phi^{-1}(p)\sigma(x') = \Big[\Phi^{-1}(g(x)) - \Phi^{-1}(p)\Big]\sigma(x')$$

is concave on $\mathbb{R}^m$ due to $\Phi^{-1}(p) > 0$ and continuously differentiable on $\mathbb{R}^m$ with gradient

$$\nabla \hat{g}(x) = \frac{\sigma(x')}{\phi(g(x))} \nabla g(x) + \left[\Phi^{-1}(g(x)) - \Phi^{-1}(p)\right] \nabla \sigma(x') \begin{pmatrix} I_m \\ 0 \end{pmatrix}.$$

Let $\bar{x} \in X$ be such that $g(\bar{x}) = p$ and $\hat{x} \in X$ be the element having the property $P(H(\hat{x})) > p$ or, equivalently, $\hat{g}(\hat{x}) > 0$. Then the concavity of $\hat{g}$ implies $\langle \nabla \hat{g}(\bar{x}), \hat{x} - \bar{x} \rangle > 0$ and, thus, $\nabla \hat{g}(\bar{x}) \notin N_X(\bar{x})$. Due to the equation $\nabla \hat{g}(\bar{x}) = \frac{\sigma(\bar{x}')}{\phi(g(\bar{x}))} \nabla g(\bar{x})$, we conclude $\nabla g(\bar{x}) \notin N_X(\bar{x})$. Hence, the constraint qualification (45) and, thus, condition (ii) of Theorem 39 are satisfied. □

For the remainder of this section we assume that the technology matrix $T(\cdot)$ is fixed, i.e., $T(\xi) \equiv T$. We will show that the constraint qualification of Corollary 42, i.e., $P(H(\hat{x})) > p$ for some $\hat{x} \in X$, implies condition (ii) of Theorem 39 for any r-concave probability distribution.
To recall the notion of r-concavity, we introduce first the generalized mean function m_r on $\mathbb{R}_+ \times \mathbb{R}_+ \times [0,1]$ for $r \in [-\infty, \infty]$ by

$$m_r(a,b;\lambda) := \begin{cases} (\lambda a^r + (1-\lambda) b^r)^{1/r} \ , \ r \in (0,\infty) \text{ or } r \in (-\infty, 0), ab > 0, \\ 0 \ , \ ab = 0, r \in (-\infty, 0), \\ a^\lambda b^{1-\lambda} \ , \ r = 0, \\ \max\{a,b\} \ , \ r = \infty, \\ \min\{a,b\} \ , \ r = -\infty. \end{cases} \tag{47}$$

A measure $P \in \mathcal{P}(\mathbb{R}^s)$ is called *r-concave* for some $r \in [-\infty, \infty]$ (cf. Prekopa (1995)) if the inequality

$$P(\lambda B_1 + (1-\lambda) B_2) \geq m_r(P(B_1), P(B_2); \lambda)$$

holds for all $\lambda \in [0,1]$ and all convex Borel subsets B_1, B_2 of $\mathbb{R}^s$ such that $\lambda B_1 + (1-\lambda) B_2$ is Borel. For $r = 0$ and $r = -\infty$, P is also called *logarithmic concave* and *quasi-concave*, respectively. Since $m_r(a,b;\lambda)$ is increasing in r if all the other variables are fixed, the sets of all r-concave probability measures are increasing if r is decreasing. It is known that $P \in \mathcal{P}(\mathbb{R}^s)$ is r-concave for some $r \in [-\infty, 1/s]$ if P has a density f_P such that

$$f_P(\lambda z + (1-\lambda)\tilde{z}) \geq m_{r(s)}(f_P(z), f_P(\tilde{z}); \lambda), \tag{48}$$

where $r(s) = r(1-rs)^{-1}$, holds for all $\lambda \in [0,1]$ and $z, \tilde{z} \in \mathbb{R}^s$. Let us mention that many multivariate probability distributions are r-concave for some $r \in$

$(-\infty, \infty]$, e.g. the uniform distribution (on some bounded convex set), the (nondegenerate) multivariate normal distribution, the Dirichlet distribution, the multivariate Student and Pareto distributions (see Prekopa (1995)).
The key observation of r-concave measures in the context of probabilistic constraints is the following one.

Lemma 43 *Let H be a multifunction from $\mathbb{R}^m$ to $\mathbb{R}^s$ with closed convex graph and P be r-concave for some $r \in [-\infty, \infty]$. Then the function $g := P(H(\cdot))$ from $\mathbb{R}^m$ to $\mathbb{R}$ has the property*

$$g(\lambda x + (1-\lambda)\tilde{x}) \geq m_r(g(x), g(\tilde{x}); \lambda)$$

for each $x, \tilde{x} \in \mathbb{R}^m$ and $\lambda \in [0,1]$.

Proof: In particular, $H(x)$ is a closed convex subset of $\mathbb{R}^s$ for any $x \in \mathbb{R}^m$. Let $x, \tilde{x} \in \mathbb{R}^m$ and $\lambda \in [0,1]$. Then the set $\lambda H(x) + (1-\lambda)H(\tilde{x})$ is also closed and convex and it holds that $\lambda H(x) + (1-\lambda)H(\tilde{x}) \subseteq H(\lambda x + (1-\lambda)\tilde{x})$. Using the r-concavity of P this implies

$$g(\lambda x + (1-\lambda)\tilde{x}) \geq m_r(P(H(x)), P(H(\tilde{x})); \lambda) = m_r(g(x), g(\tilde{x}); \lambda). \qquad \square$$

Corollary 44 *Let $T(\xi) \equiv T$ and P be r-concave for some $r \in (-\infty, \infty]$. Let $X^*(P)$ be nonempty and $\mathcal{U} \subset \mathbb{R}^m$ be an open bounded neighbourhood of $X^*(P)$. Assume that there exists an element $\hat{x} \in X$ such that $P(H(\hat{x})) > p$ holds. Then there are constants $L > 0$, $\delta > 0$ and $k \in \mathbb{N}$ such that*

$$\begin{aligned} |\vartheta(P) - \vartheta_{\mathcal{U}}(Q)| &\leq L\, \alpha_{\mathrm{ph}_k}(P,Q) \\ \emptyset \neq X^*_{\mathcal{U}}(Q) &\subseteq X^*(P) + \Psi_P(L\alpha_{\mathrm{ph}_k}(P,Q))\mathbb{B}, \end{aligned}$$

and $X^*_{\mathcal{U}}(Q)$ is a CLM set for (44) relative to $\mathcal{U}$ whenever $Q \in \mathcal{P}(\Xi)$ and $\alpha_{\mathrm{ph}_k}(P,Q) < \delta$. Here, the function Ψ_P is given by (23).

Proof: We assume without loss of generality that $r < 0$. Again we have to verify the metric regularity condition (ii) of Theorem 39. To this end, we use the function $\hat{g}(\cdot) := p^r - g^r(\cdot)$ instead of $g(\cdot) := P(H(\cdot))$. Since P is r-concave, the function $\hat{g}(\cdot)$ is concave on $\mathbb{R}^m$. We consider the set-valued mapping $\Gamma(x) := \{v \in \mathbb{R} : x \in X, \hat{g}(x) \geq v\}$ from $\mathbb{R}^m$ to $\mathbb{R}$. Its graph is closed and convex. Let $\bar{x} \in X$ with $g(\bar{x}) = p$, i.e., $\hat{g}(\bar{x}) = p^r$. As there exists an $\hat{x} \in X$ such that $g(\hat{x}) > p$, i.e., $\hat{g}(\hat{x}) > 0$, the element $v = 0$ belongs to the interior of the range of Γ. Hence, the Robinson-Ursescu Theorem (Theorem 9.48 in Rockafellar and Wets (1998)) implies the existence of constants $a > 0$ and $\varepsilon > 0$ such that

$$d(x, \Gamma^{-1}(v)) \leq a d(v, \Gamma(x)) \leq a \max\{0, v - \hat{g}(x)\}$$

holds whenever $x \in X$, $\|x - \bar{x}\| \leq \varepsilon$ and $|v| \leq \varepsilon$. For $x \in X$ with $\|x - \bar{x}\| \leq \varepsilon$ and sufficiently small $|y|$ we obtain

$$d(x, \mathcal{X}_y(P)) = d(x, \Gamma^{-1}(p^r - (p-y)^r)) \leq a \max\{0, g^r(x) - (p-y)^r\}$$

Finally, it remains to use that the function $v \mapsto v^r$ is locally Lipschitz continuous on $(0, +\infty)$. $\square$

The above result improves in case of $h(\xi) \equiv \xi$ and, hence, $g(x) = F_P(Tx)$, where F_P is the distribution function of P. Then the polyhedral discrepancy α_{ph_k} can be replaced by the Kolmogorov distance d_{K}.
The next result provides a sufficient condition for (ii) in situations where P is not quasiconcave, but has a density on $\mathbb{R}^s$. Here, metric regularity is implied by a growth condition of $g(\cdot) = F_P(T\cdot)$ (see Henrion and Römisch (1999)).

Corollary 45 *Let $T(\xi) \equiv T$, $h(\xi) \equiv \xi$, $P \in \mathcal{P}(\mathbb{R}^s)$ have a density f_P, $X^*(P)$ be nonempty and $\mathcal{U} \subseteq \mathbb{R}^m$ be an open bounded neighbourhood of $X^*(P)$. Assume the following two conditions for each $\bar{x} \in X^*(P)$:*

(i) $(T\bar{x} + \mathrm{bd}\,\mathbb{R}^s_-) \cap \{\xi \in \mathbb{R}^s : \exists \varepsilon > 0$ *such that* $f_P(\eta) \geq \varepsilon, \forall \eta \in \xi + \varepsilon\mathbb{B}\} \neq \emptyset$,
(ii) there exists an $\hat{x} \in X$ such that $T\hat{x} > T\bar{x}$ holds componentwise.

Then there are constants $L > 0$ and $\delta > 0$ such that

$$|\vartheta(P) - \vartheta_{\mathcal{U}}(Q)| \leq L\, d_{\mathrm{K}}(P, Q)$$
$$\emptyset \neq X^*_{\mathcal{U}}(Q) \subseteq X^*(P) + \Psi_P(L\, d_{\mathrm{K}}(P, Q))\mathbb{B},$$

and $X^*_{\mathcal{U}}(Q)$ is a CLM set of (44) relative to $\mathcal{U}$ whenever $Q \in \mathcal{P}(\Xi)$ and $d_{\mathrm{K}}(P, Q) < \delta$. Here, the function Ψ_P is given by (23).

The essential condition (i) says that, for each $\chi \in T(X^*(P))$, the boundary of the cell $\chi + \mathbb{R}^s_-$ meets the strict positivity region of the density of P somewhere. This implies a suitable growth behaviour of the distribution function F_P at elements of $T(X^*(P))$ and, hence, metric regularity.
Finally, we study the growth function ψ_P of (44) and derive conditions implying quadratic growth near solution sets in case of $h(\xi) \equiv \xi$ and a logarithmic concave measure P. The first step of our analysis consists in a reduction argument that decomposes problem (44) into two auxiliary problems. The first one is a stochastic program with modified objective and probabilistic constraints (with decisions taken in $\mathbb{R}^s$) whereas the second one represents a parametric linear program. The argument is similar to Lemma 28 for two-stage models and was proved in Henrion and Römisch (1999).

Lemma 46 *Let $Q \in \mathcal{P}(\mathbb{R}^s)$ and $\mathcal{U} \subseteq \mathbb{R}^m$ be a nonempty open set such that* *its closure is a polytope. Then we have*

$$\vartheta_{\mathcal{U}}(Q) = \inf\{\pi_{\mathcal{U}}(y) : y \in T(X_{\mathcal{U}}), F_Q(y) \geq p\} \quad \textit{and} \quad X^*_{\mathcal{U}}(Q) = \sigma_{\mathcal{U}}(Y_{\mathcal{U}}(Q)),$$

where

$$\begin{aligned} X_{\mathcal{U}} &= X \cap \operatorname{cl}\mathcal{U}, \\ Y_{\mathcal{U}}(Q) &= \operatorname{argmin}\{\pi_{\mathcal{U}}(y) : y \in T(X_{\mathcal{U}}), F_Q(y) \geq p\}, \\ \pi_{\mathcal{U}}(y) &= \inf\{\langle c, x\rangle : Tx = y, x \in X_{\mathcal{U}}\}, \\ \sigma_{\mathcal{U}}(y) &= \operatorname{argmin}\{\langle c, x\rangle : Tx = y, x \in X_{\mathcal{U}}\} \ (y \in T(X_{\mathcal{U}})). \end{aligned}$$

Here, $\pi_{\mathcal{U}}$ is convex polyhedral on $T(X_{\mathcal{U}})$ and $\sigma_{\mathcal{U}}$ is Lipschitz continuous on $T(X_{\mathcal{U}})$ with respect to the Pompeiu-Hausdorff distance on $\mathbb{R}^s$.

Theorem 47 *Let $T(\xi) \equiv T$, $h(\xi) \equiv \xi$, $P \in \mathcal{P}(\mathbb{R}^s)$ be logarithmic concave and $X^*(P)$ be nonempty and bounded. Assume that*

(i) $X^*(P) \cap \operatorname{argmin}\{\langle c, x\rangle : x \in X\} = \emptyset$;
(ii) there exists an $\bar{x} \in X$ such that $F_P(T\bar{x}) > p$;
(iii) $\log F_P$ *is strongly concave on some convex neighbourhood $\mathcal{V}$ of $T(X^*(P))$.*

Then there exist $L > 0$ and $\delta > 0$ and a neighbourhood $\mathcal{U}$ of $X^*(P)$ such that

$$\mathbb{D}_\infty(X^*(P), X^*_{\mathcal{U}}(Q)) \leq L d_{\mathrm{K}}(P, Q)^{1/2}$$

holds whenever $Q \in \mathcal{P}(\mathbb{R}^s)$ and $d_{\mathrm{K}}(P, Q) < \delta$. Here, $\mathbb{D}_\infty$ denotes the Pompeiu-Hausdorff distance on subsets of $\mathbb{R}^m$ and d_{K} the Kolmogorov metric on $\mathcal{P}(\mathbb{R}^s)$.

Proof: Let $\mathcal{U}_0 \subseteq \mathbb{R}^m$ be an open convex set such that $X^*(P) \subseteq \mathcal{U}_0$ and $T(\mathcal{U}_0) \subseteq \mathcal{V}$. For each $x \in X^*(P)$ select $\varepsilon(x) > 0$ such that the polyhedron $x + \varepsilon(x)\mathbb{B}_\infty$ (with $\mathbb{B}_\infty$ denoting the closed unit ball w.r.t. the norm $\|\cdot\|_\infty$ on $\mathbb{R}^m$) is contained in $\mathcal{U}_0$. Since $X^*(P)$ is compact, finitly many of these balls cover $X^*(P)$. The closed convex hull $\bar{\mathcal{U}}$ of their union is a polyhedron with $X^*(P) \subseteq \mathcal{U} \subset \bar{\mathcal{U}} \subseteq \mathcal{U}_0$, where $\mathcal{U} = \operatorname{int} \bar{\mathcal{U}}$. With the notations of Lemma 46 we consider the problem

$$\min\{\pi_{\mathcal{U}}(y) : y \in T(X_{\mathcal{U}}), \hat{g}(y) := \log p - \log F_P(y) \leq 0\}.$$

According to Lemma 46 the solution set $Y_{\mathcal{U}}(P)$ of this problem fulfils $X^*(P) = X^*_{\mathcal{U}}(P) = \sigma_{\mathcal{U}}(Y_{\mathcal{U}}(P))$. Let $y_* \in Y_{\mathcal{U}}(P)$ and $\bar{y} = T\bar{x}$ with $\bar{x} \in X$ from (ii). Then the logarithmic concavity of P implies for any $\lambda \in (0, 1]$:

$$\begin{aligned} \hat{g}(\lambda\bar{y} + (1-\lambda)y_*) &= \log p - \log F_P(\lambda\bar{y} + (1-\lambda)y_*) \\ &\leq \log p - \lambda \log F_P(\bar{y}) - (1-\lambda)\log F_P(y_*) \\ &\leq \lambda(\log p - \log F_P(\bar{y})) < 0. \end{aligned}$$

Thus, we may choose $\hat{\lambda} \in (0,1]$ such that $\hat{y} = \hat{\lambda}\bar{y} + (1-\hat{\lambda})y_*$ belongs to $T(X_\mathcal{U})$ and has the property $\hat{g}(\hat{y}) < 0$. This constraint qualification implies the existence of a Kuhn-Tucker coefficient $\lambda_* \geq 0$ such that

$$\pi_\mathcal{U}(y_*) = \min\{\pi_\mathcal{U}(y) + \lambda_*\hat{g}(y) : y \in T(X_\mathcal{U})\} \quad \text{and} \quad \lambda_*\hat{g}(y_*) = 0.$$

In case $\lambda_* = 0$, this would imply $y_* \in \operatorname{argmin}\{\pi_\mathcal{U}(y) : y \in T(X_\mathcal{U})\}$ and, hence, the existence of some $x_* \in X^*(P)$ with $\langle c, x_*\rangle = \pi_\mathcal{U}(Tx_*) = \min\{\langle c, x\rangle : Tx = y_*, x \in X_\mathcal{U}\}$. Hence, condition (i) would be violated due to $x^* \in \operatorname{int}\mathcal{U}$. Thus $\lambda_* > 0$ and $\pi_V + \lambda_*\hat{g}$ is strongly convex on $T(X_\mathcal{U})$. Hence, y_* is the unique minimizer of $\pi_V + \lambda_*\hat{g}$ and the growth property

$$\rho\|y - y_*\|^2 \leq \pi_\mathcal{U}(y) + \lambda_*\hat{g}(y) - \pi_\mathcal{U}(y^*) \tag{49}$$

holds for some $\rho > 0$ and all $y \in T(X_\mathcal{U})$.
As the assumptions of Corollary 44 are satisfied, the set-valued mapping $X^*_\mathcal{U}(\cdot)$ is upper semicontinuous at P and $X^*_\mathcal{U}(Q) \neq \emptyset$ is a complete local minimizing set if $d_\mathrm{K}(P,Q)$ is sufficiently small. Hence, there exists a $\delta > 0$ such that $\emptyset \neq X^*_\mathcal{U}(Q) \subseteq \mathcal{U}$ for all $Q \in \mathcal{P}(\mathbb{R}^s)$ with $d_\mathrm{K}(P.Q) < \delta$. With the notations from Lemma 46 and using the fact that $Y_\mathcal{U}(P) = \{y_*\}$ and $X^*(P) = X^*_\mathcal{U}(P) = \sigma_\mathcal{U}(y_*)$ we obtain

$$\mathbb{D}_\infty(X^*(P), X^*_\mathcal{U}(Q)) = \mathbb{D}_\infty(\sigma_\mathcal{U}(y_*), \sigma_\mathcal{U}(Y_\mathcal{U}(Q))) \leq \hat{L} \sup_{y \in Y_\mathcal{U}(Q)} \|y - y_*\|,$$

where $\hat{L} > 0$ is the Lipschitz constant of $\sigma_\mathcal{U}$ (cf. Lemma 46). Using (49) and $Y_\mathcal{U}(Q) \subseteq T(X_\mathcal{U})$, the above chain of inequalities extends to

$$\begin{aligned}
\mathbb{D}_\infty(X^*(P), X^*_\mathcal{U}(Q)) &\leq \frac{\hat{L}}{\rho^{1/2}} \sup_{y \in Y_\mathcal{U}(Q)} [\pi_\mathcal{U}(y) + \lambda_*\hat{g}(y) - \pi_\mathcal{U}(y_*)]^{1/2} \\
&= \frac{\hat{L}}{\rho^{1/2}}[\vartheta_\mathcal{U}(Q) - \vartheta(P) + \lambda_*(\log p - \log F_P(y))]^{1/2} \\
&\leq \frac{\hat{L}}{\rho^{1/2}}[\vartheta_\mathcal{U}(Q) - \vartheta(P) + \lambda_*(\log F_Q(y) - \log F_P(y))]^{1/2} \\
&\leq \frac{\hat{L}}{\rho^{1/2}}[(L + \frac{\lambda_*}{p})d_\mathrm{K}(P,Q)]^{1/2},
\end{aligned}$$

where $L > 0$ is the constant from Theorem 39 and $\frac{1}{p}$ the Lipschitz constant of $\log(\cdot)$ on $[p, 1]$. This completes the proof. □

A slightly more general version of the result for r-concave measures was proved in Henrion and Römisch (1999). The assumptions (i)–(iii) imposed in Theorem 47 concern the original problem. The conditions (i) and (ii) mean that the probability level p is not chosen too low and too high, respectively. Condition (i) expresses the fact that the presence of the probabilistic constraint $F_P(Tx) \geq p$ moves the solution set $X^*(P)$ away from the one obtained without imposing that constraint. Recent results in Henrion and Römisch (2002) show that assumption (i) is not necessary for Theorem 47 to hold. Assumption (iii) is decisive for the desired growth condition of the objective function around $X^*(P)$. In contrast to the global concavity of $\log F_P$, (iii) requires the strong concavity of $\log F_P$ as a local property around $T(X^*(P))$. Since general sufficient criteria for (iii) are not available so far, we provide a few examples.

Example 48 (strong logarithmic concavity of measures)
Let P be the uniform distribution on some bounded rectangle in $\mathbb{R}^s$ having the form $D = \times_{i=1}^s [a_i, b_i]$. Then $\log F_P(\xi) = \sum_{i=1}^s \log(\xi_i - a_i)$, $\xi \in D$. Clearly, $\log(\cdot - a_i)$ is strongly concave on any closed subinterval of (a_i, b_i). Hence, $\log F_P(\cdot)$ is strongly concave on any closed convex subset of $\operatorname{int} D$.
Let P be the multivariate normal distribution on $\mathbb{R}^s$ having a nonsingular diagonal covariance matrix. A direct computation for the standard normal distribution function Φ on $\mathbb{R}$ shows that $\log \Phi$ is strongly concave on any bounded interval. Since $\log F_P$ is equal to the sum of logarithms of the marginal distribution functions, it is strongly concave on any bounded convex set in $\mathbb{R}^s$.

4 Approximations of Stochastic Programs

Many approximations of stochastic programs result from replacing the underlying probability distribution by some other measure, which typically leads to simpler models. Important examples are nonparametric statistical estimates (e.g. empirical ones) and scenario tree constructions using probability distribution information. Next we give an idea how the results of the previous sections may be used to design and to analyse approximations of stochastic programs. We begin with some glimpses into the analysis of empirical approximations and the relations to empirical process theory. A more far-reaching analysis is given in Pflug (2003) and Shapiro (2003).

4.1 A Glimpse of Empirical Approximations

Let $P \in \mathcal{P}(\Xi)$ and $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ be independent identically distributed Ξ-valued random variables on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ having the common distribution P, i.e., $P = \mathbb{P}\xi_1^{-1}$. We consider the *empirical measures*

$$P_n(\omega) := \frac{1}{n} \sum_{i=1}^n \delta_{\xi_i(\omega)} \qquad (\omega \in \Omega;\ n \in \mathbb{N}),$$

where δ_ξ denotes the unit mass at $\xi \in \Xi$, and the empirical approximations of the stochastic program (1), i.e., the models that result from replacing P by $P_n(\cdot)$. These take the form

$$\min\Big\{\sum_{i=1}^{n} F_0(x,\xi_i(\cdot)) \,:\, x\in X\,,\,\sum_{i=1}^{n} F_j(x,\xi_i(\cdot))\le 0\,,\, j=1,\ldots,d\Big\}, \qquad (50)$$

where the factor $\frac{1}{n}$ in the objective and constraints has been removed. Since the objective and constraint functions F_j, $j=0,\ldots,d$, are assumed to be random lower semicontinuous functions from $\mathbb{R}^m\times\Xi$ to $\overline{\mathbb{R}}$, the constraint set is closed-valued and measurable from Ω to $\mathbb{R}^m$ and, hence, the optimal value $\vartheta(P_n(\cdot))$ of (50) is measurable from Ω to $\overline{\mathbb{R}}$ and the solution set $X^*(P_n(\cdot))$ is a closed-valued measurable multifunction from Ω to $\mathbb{R}^m$ (see Chapter 14 and, in particular, Theorem 14.37 in Rockafellar and Wets (1998)). The same conclusion is valid for the localized concepts $\vartheta_{\mathcal{U}}$ and $X^*_{\mathcal{U}}$ for any nonempty open subset $\mathcal{U}$ of $\mathbb{R}^m$.

Another measurability question arises when studying uniform convergence properties of the *empirical process*

$$\Big\{n^{\frac{1}{2}}(P_n(\cdot)-P)F = n^{-\frac{1}{2}}\sum_{i=1}^{n}(F(\xi_i(\cdot))-PF)\Big\}_{F\in\mathcal{F}},$$

indexed by some class $\mathcal{F}$ of functions that are integrable with respect to P. Here, we set $QF := \int_\Xi F(\xi)dQ(\xi)$ for any $Q\in\mathcal{P}(\Xi)$ and $F\in\mathcal{F}$. Since the suprema $d_{\mathcal{F}}(P_n(\cdot),P)=\sup_{F\in\mathcal{F}}|P_n(\cdot)F-PF|$ may be non-measurable functions from Ω to $\overline{\mathbb{R}}$, we introduce a condition on $\mathcal{F}$ that simplifies matters and is satisfied in most stochastic programming models. A class $\mathcal{F}$ of measurable functions from Ξ to $\overline{\mathbb{R}}$ is called *P-permissible* for some $P\in\mathcal{P}(\Xi)$ if there exists a countable subset $\mathcal{F}_0$ of $\mathcal{F}$ such that for each function $F\in\mathcal{F}$ there exists a sequence (F_n) in $\mathcal{F}_0$ converging pointwise to F and such that the sequence (PF_n) also converges to PF. Then

$$d_{\mathcal{F}}(P_n(\omega),P)=\sup_{F\in\mathcal{F}}|(P_n(\omega)-P)F| = d_{\mathcal{F}_0}(P_n(\omega),P)$$

holds for each $n\in\mathbb{N}$ and $\omega\in\Omega$, i.e., the analysis is reduced to a countable class and, in particular, $d_{\mathcal{F}}(P_n(\cdot),P)$ is a measurable function from Ω to $\overline{\mathbb{R}}$.

A P-permissible class $\mathcal{F}$ is called a *P-Glivenko-Cantelli class* if the sequence $(d_{\mathcal{F}}(P_n(\cdot),P))$ of random variables converges to 0 $\mathbb{P}$-almost surely. If $\mathcal{F}$ is P-permissible, the empirical process $\{n^{\frac{1}{2}}(P_n(\cdot)-P)F\}_{F\in\mathcal{F}}$ is called *uniformly bounded in probability with tail* $C_{\mathcal{F}}(\cdot)$ if the function $C_{\mathcal{F}}(\cdot)$ is defined on $(0,\infty)$ and decreasing to 0, and the estimate

$$\mathbb{P}(\{\omega : n^{\frac{1}{2}}d_{\mathcal{F}}(P_n(\omega),P)\ge\varepsilon\})\le C_{\mathcal{F}}(\varepsilon) \qquad (51)$$

holds for each $\varepsilon > 0$ and $n \in \mathbb{N}$. Whether a given class $\mathcal{F}$ is a P-Glivenko-Cantelli class or the empirical process is uniformly bounded in probability depends on the size of the class $\mathcal{F}$ measured in terms of certain *covering numbers* or the corresponding *metric entropy numbers* defined as their logarithms (e.g., Dudley (1984), Pollard (1990), van der Vaart and Wellner (1996)). To introduce these concepts, let $\mathcal{F}$ be a subset of the normed space $L_r(\Xi, P)$ for some $r \geq 1$ equipped with the usual norm $\|F\|_{P,r} = (P|F|^r)^{\frac{1}{r}}$. The covering number $N(\varepsilon, \mathcal{F}, L_r(\Xi, P))$ is the minimal number of open balls $\{G \in L_r(\Xi, P) : \|G - F\|_{P,r} < \varepsilon\}$ needed to cover $\mathcal{F}$. A measurable function $F_{\mathcal{F}}$ from Ξ to $\overline{\mathbb{R}}$ is called an *envelope* of the class $\mathcal{F}$ if $|F(\xi)| \leq F_{\mathcal{F}}(\xi)$ holds for every $\xi \in \Xi$ and $F \in \mathcal{F}$. The following result provides criteria for the desired properties in terms of uniform covering numbers.

Theorem 49 *Let $\mathcal{F}$ be P-permissible with envelope $F_{\mathcal{F}}$. If $PF_{\mathcal{F}} < \infty$ and*

$$\sup_Q N(\varepsilon \|F_{\mathcal{F}}\|_{Q,1}, \mathcal{F}, L_1(Q)) < \infty, \tag{52}$$

then $\mathcal{F}$ is a P-Glivenko-Cantelli class. If $\mathcal{F}$ is uniformly bounded and there exist constants $r \geq 1$ and $R \geq 1$ such that

$$\sup_Q N(\varepsilon \|F_{\mathcal{F}}\|_{Q,2}, \mathcal{F}, L_2(Q)) \leq \Big(\frac{R}{\varepsilon}\Big)^r \tag{53}$$

holds for all $\varepsilon > 0$, then the empirical process indexed by $\mathcal{F}$ is uniformly bounded in probability with exponential tail $C_{\mathcal{F}}(\varepsilon) = (K(R)\varepsilon r^{-\frac{1}{2}})^r \exp(-2\varepsilon^2)$ with some constant $K(R)$ depending only on R.
The suprema in (52) and (53) are taken over all finitely discrete probability measures Q with $\|F_{\mathcal{F}}\|_{Q,1} = QF_{\mathcal{F}} > 0$ and $\|F_{\mathcal{F}}\|^2_{Q,2} = QF^2_{\mathcal{F}} > 0$, respectively.

For the proof we refer to Talagrand (1994), van der Vaart and Wellner (1996) and van der Vaart (1998). For studying entropic sizes of stochastic programs Pflug (1999, 2003) uses results of this type but with *bracketing numbers* instead of uniform covering numbers. He also studies situations where $\mathcal{F}$ is not uniformly bounded and shows that the *blow-up function* $n^{\frac{1}{2}}$ for $n \to \infty$ has to be replaced by some function converging to ∞ more slowly. Here, we use the concept of uniform covering numbers since they turn out to be useful for discontinuous functions.
The stability results of Section 2 directly translate into convergence results and rates, respectively, for empirical optimal values and solution sets.

Theorem 50 *Assume that the conditions (i)–(iii) of Theorem 5 are satisfied and that $\mathcal{F}_{\mathcal{U}}$ is P-permissible.*

If $\mathcal{F}_{\mathcal{U}}$ is a P-Glivenko-Cantelli class, the sequences

$$\Big(|\vartheta(P) - \vartheta_{\mathcal{U}}(P_n(\cdot))|\Big) \quad and \quad \Big(\sup_{x \in X^*_{\mathcal{U}}(P_n(\cdot))} d(x, X^*(P))\Big)$$

converge $\mathbb{P}$-almost surely to 0. Furthermore, the set $X^*_{\mathcal{U}}(P_n(\omega))$ is a CLM set of (50) relative to $\mathcal{U}$ for sufficiently large $n \in \mathbb{N}$ and for $\mathbb{P}$-almost all $\omega \in \Omega$. If the empirical process indexed by $\mathcal{F}_{\mathcal{U}}$ is uniformly bounded in probability with tail $C_{\mathcal{F}_{\mathcal{U}}}(\cdot)$, the following estimates hold for each $\varepsilon > 0$ and each $n \in \mathbb{N}$:

$$\mathbb{P}(|\vartheta(P) - \vartheta_{\mathcal{U}}(P_n(\cdot))| > \varepsilon n^{-\frac{1}{2}}) \leq C_{\mathcal{F}_{\mathcal{U}}}(\min\{\delta, \frac{\varepsilon}{L}\}), \tag{54}$$

$$\mathbb{P}(\sup_{x \in X^*_{\mathcal{U}}(P_n(\cdot))} d(x, X^*(P)) > \varepsilon n^{-\frac{1}{2}}) \leq C_{\mathcal{F}_{\mathcal{U}}}(\min\{\delta, \hat{L}^{-1}\Psi_P^{-1}(\varepsilon)\}). \tag{55}$$

Proof: Let $L > 0$, $\hat{L} > 0$, $\delta > 0$ be the constants in Theorems 5 and 9. First, let $\mathcal{F}_{\mathcal{U}}$ be a P-Glivenko-Cantelli class and $A \in \mathcal{A}$ be such that $\mathbb{P}(A) = 0$ and $(d_{\mathcal{F}_{\mathcal{U}}}(P_n(\omega), P))$ converges to 0 for every $\omega \in \Omega \setminus A$. Let $\omega \in \Omega \setminus A$. Then $X^*_{\mathcal{U}}(P_n(\omega))$ is nonempty, since the objective function $\int_\Xi F_0(\cdot, \xi) dP(\xi)$ is lower semicontinuous on X and the constraint set $\mathcal{X}_{\mathcal{U}}(P_n(\omega))$ is compact due to Proposition 3. Let $n_0(\omega) \in \mathbb{N}$ be such that $d_{\mathcal{F}_{\mathcal{U}}}(P_n(\omega), P) < \delta$ holds for each $n \geq n_0(\omega)$. Due to the Theorems 5 and 9 the estimates

$$\begin{aligned} |\vartheta(P) - \vartheta_{\mathcal{U}}(P_n(\omega))| &\leq L d_{\mathcal{F}_{\mathcal{U}}}(P_n(\omega), P) \\ \sup_{x \in X^*_{\mathcal{U}}(P_n(\omega))} d(x, X^*(P)) &\leq \Psi_P(\hat{L} d_{\mathcal{F}_{\mathcal{U}}}(P_n(\omega), P)) \end{aligned}$$

hold for $n \geq n_0(\omega)$. In particular, the sequences $(|\vartheta(P) - \vartheta_{\mathcal{U}}(P_n(\omega))|)$ and $(\sup_{x \in X^*_{\mathcal{U}}(P_n(\omega))} d(x, X^*(P)))$ converge to 0. Hence, $X^*_{\mathcal{U}}(P_n(\omega)) \subseteq \mathcal{U}$ and, thus, $X^*_{\mathcal{U}}(P_n(\omega))$ is a CLM set relative to $\mathcal{U}$ for sufficiently large $n \in \mathbb{N}$.
Now, let $\varepsilon > 0$ be arbitrary. The Theorems 5 and 9 also imply

$$\mathbb{P}(|\vartheta(P) - \vartheta_{\mathcal{U}}(P_n(\cdot))| > \varepsilon) \leq \mathbb{P}(d_{\mathcal{F}_{\mathcal{U}}}(P_n(\cdot), P) \geq \min\{\delta, \frac{\varepsilon}{L}\}), \tag{56}$$

$$\mathbb{P}(\sup_{x \in X^*_{\mathcal{U}}(P_n(\cdot))} d(x, X^*(P)) > \varepsilon) \leq \mathbb{P}(d_{\mathcal{F}_{\mathcal{U}}}(P_n(\cdot), P) \geq \min\{\delta, \hat{L}^{-1}\Psi_P^{-1}(\varepsilon)\}). \tag{57}$$

If the empirical process indexed by $\mathcal{F}_{\mathcal{U}}$ is uniformly bounded in probability with tail $C_{\mathcal{F}_{\mathcal{U}}}(\cdot)$, the estimates (56) and (57) may be continued by using (51) and, thus, lead to (54) and (55). $\Box$

The estimates (54) and (55) may be used to derive the speed of convergence in probability of optimal values and solution sets, respectively. Clearly, the

speed depends on the asymptotic behaviour of the tail $C_{\mathcal{F}_{\mathcal{U}}}(\varepsilon)$ as $\varepsilon \to \infty$ and of the function Ψ_P. For the situation of exponential tails, this is elaborated in Rachev and Römisch (2002).
Next we show how our analysis applies to two-stage stochastic programs with and without integrality requirements and to chance constrained models. It turns out that, under reasonable assumptions on all models, the empirical process indexed by $\mathcal{F}_{\mathcal{U}}$ is uniformly bounded in probability with exponential tails.

Example 51 (linear chance constrained models)
A class $\mathcal{B}$ of Borel sets of $\mathbb{R}^s$ is called a *Vapnik-Červonenkis (VC) class* of index $r = r(\mathcal{B})$ if r is finite and equal to the smallest $n \in \mathbb{N}$ for which no set of cardinality $n+1$ is shattered by $\mathcal{B}$. $\mathcal{B}$ is said to shatter a subset $\{\xi_1, \ldots, \xi_l\}$ of cardinality l in $\mathbb{R}^s$ if each of its 2^l subsets is of the form $B \cap \{\xi_1, \ldots, \xi_l\}$ for some $B \in \mathcal{B}$. For VC classes $\mathcal{B}$ it holds that

$$N(\varepsilon, \{\chi_B : B \in \mathcal{B}\}, L_1(\Xi, Q)) \leq K\varepsilon^{-r}$$

for any $\varepsilon > 0$ and $Q \in \mathcal{P}(\Xi)$, and some constant $K > 0$ depending only on the index r (Theorem 2.6.4 in van der Vaart and Wellner (1996)).
For any polyhedral set $\Xi \subseteq \mathbb{R}^s$ and $k \in \mathbb{N}$ the class $\mathcal{B}_{\mathrm{ph}_k}(\Xi)$ is a VC class, since the class of all closed half spaces is VC and finite intersections of VC classes are again VC. The corresponding class of characteristic functions is permissible for P, since the set of all polyhedra in $\mathcal{B}_{\mathrm{ph}_k}(\Xi)$ having vertices at rational points in $\mathbb{R}^s$ plays the role of the countable subset in the definition of permissibility. Hence, Theorem 49 applies and the empirical process indexed by $\mathcal{F}_{\mathcal{U}} = \{\chi_{H(x)} : x \in X \cap \mathrm{cl}\,\mathcal{U}\}$, where $\mathcal{U}$ is a bounded open set containing $X^*(P)$, is uniformly bounded in probability with exponential tail $C_{\mathcal{F}_{\mathcal{U}}}(\varepsilon) = \hat{K}\varepsilon^r \exp(-2\varepsilon^2)$ for some index $r \in \mathbb{N}$ and some constant $\hat{K} > 0$. For example, from Theorem 50 we obtain for each $\varepsilon > 0$ and $n \in \mathbb{N}$ the estimate

$$\mathbb{P}\Big(\sup_{x \in X^*_{\mathcal{U}}(P_n(\cdot))} d(x, X^*(P)) > \varepsilon n^{-\frac{1}{2}}\Big) \leq \hat{K}\varepsilon^r \exp(-2\min\{\delta, \hat{L}^{-1}\Psi_P^{-1}(\varepsilon)\}^2).$$

Example 52 (two-stage models without integrality)
Let F_0 be defined as in Section 3.1 and let (A1) and (A2) be satisfied. Then, for each nonempty open and bounded subset $\mathcal{U}$ of $\mathbb{R}^m$, the class $\mathcal{F}_{\mathcal{U}} = \{F_0(x, \cdot) : x \in X \cap \mathrm{cl}\,\mathcal{U}\}$ is a subset of $L_1(\Xi, P)$. $\mathcal{F}_{\mathcal{U}}$ is also permissible for P, since any class $\{F_0(x, \cdot) : x \in X_c\}$ with X_c being a countable and dense subset of $X \cap \mathrm{cl}\,\mathcal{U}$ may be used as the countable subset of $\mathcal{F}_{\mathcal{U}}$ in the definition of permissibility. Proposition 22 implies that the function $F_{\mathcal{F}_{\mathcal{U}}}(\xi) := K\max\{1, \|\xi\|^2\}$ ($\xi \in \Xi$) is an envelope of $\mathcal{F}_{\mathcal{U}}$ for sufficiently large $K > 0$. Furthermore, due to the Lipschitz continuity property of $F_0(\cdot, \xi)$ with Lipschitz constant $\hat{L}\max\{1, \|\xi\|^2\}$ (see Proposition 22), the uniform covering numbers of $\mathcal{F}_{\mathcal{U}}$ are bounded by the covering numbers of $X \cap \mathrm{cl}\,\mathcal{U}$ (see Theorem 2.7.11 in van der Vaart and Wellner (1996)). In particular, for each finitely discrete measure

$Q \in \mathcal{P}(\Xi)$ and with $\hat{F}(\xi) := \hat{L}\max\{1, \|\xi\|^2\}$ ($\xi \in \Xi$) it holds that

$$N(\varepsilon\|\hat{F}\|_{Q,r}, \mathcal{F}_{\mathcal{U}}, L_r(\Xi, Q)) \leq N(\varepsilon, X \cap \mathrm{cl}\mathcal{U}, \mathbb{R}^m) \leq K\varepsilon^{-m}, \qquad (58)$$

for each $\varepsilon > 0$, $r \geq 1$ and some constant $K > 0$ depending only on m and the diameter of $X \cap \mathrm{cl}\mathcal{U}$. Using (58) for $r = 1$, Theorem 49 implies that $\mathcal{F}_{\mathcal{U}}$ is a P-Glivenko-Cantelli class. If Ξ is bounded, $\mathcal{F}_{\mathcal{U}}$ is uniformly bounded and, using (58) for $r = 2$, Theorem 49 implies that the empirical process indexed by $\mathcal{F}_{\mathcal{U}}$ is uniformly bounded in probability with exponential tail.

Example 53 (mixed-integer two-stage models)
Let F_0 be defined as in Section 3.2 and let (B1)–(B3) be satisfied and Ξ be bounded. Then, for each nonempty open and bounded subset $\mathcal{U}$ of $\mathbb{R}^m$, the class

$$\mathcal{F}_{\mathcal{U}} = \Big\{F_0(x,\cdot) = \sum_{j=1}^{\nu}(\langle c, x\rangle + \Phi(h(\cdot) - T(\cdot)x)\chi_{\Xi^R_{j,x}}(\cdot) : x \in X \cap \mathrm{cl}\mathcal{U}\Big\}$$

is a subset of $L_1(\Xi, P)$. This representation follows from Proposition 34 if $R > 0$ is chosen sufficiently large such that $\{\xi \in \Xi : \|h(\xi) - T(\xi)x\|_\infty > R\} = \emptyset$ for each $x \in X \cap \mathrm{cl}\mathcal{U}$. For each $X \cap \mathrm{cl}\mathcal{U}$ the sets $\Xi^R_{j,x}$ ($j = 1\ldots,\nu$) form a disjoint partition of Ξ into Borel sets whose closures are in $\mathcal{B}_{\mathrm{ph}_k}(\Xi)$ for some $k \in \mathbb{N}$. Furthermore, the function $\Phi(h(\cdot) - T(\cdot)x)$ is Lipschitz continuous on each of these sets with a uniform constant $L_1 > 0$. Let $F_0^j(x,\cdot)$ denote a Lipschitz extension of the function $\langle c, x\rangle + \Phi(h(\cdot) - T(\cdot)x)$ from $\Xi^R_{j,x}$ to $\mathbb{R}$ by preserving the Lipschitz constant L_1 ($j = 1, \ldots, \nu$). Furthermore, let $\mathcal{F}^j_{\mathcal{U}} := \{F_0^j(x,\cdot) : x \in X \cap \mathrm{cl}\mathcal{U}\}$ and $\mathcal{G}^j_{\mathcal{U}} := \{\chi_{\Xi^R_{j,x}} : x \in X \cap \mathrm{cl}\mathcal{U}\}$ ($j = 1, \ldots, \nu$). Now, we use a permanence property of the uniform covering numbers (cf. Section 2.10.3 in van der Vaart and Wellner (1996)). Let $Q \in \mathcal{P}(\Xi)$ be discrete with finite support. Then the estimate

$$N(\varepsilon C_0, \mathcal{F}_{\mathcal{U}}, L_2(\Xi, Q)) \leq \prod_{j=1}^{\nu} N(\varepsilon C_j, \mathcal{F}^j_{\mathcal{U}}, L_2(\Xi, Q_j))N(\varepsilon\hat{C}_j, \mathcal{G}^j_{\mathcal{U}}, L_2(\Xi, \hat{Q}_j)) \quad (59)$$

is valid, where C_0, $C_j > 1$, $\hat{C}_j$, $j = 1, \ldots, \nu$, are certain constants and Q_j, $\hat{Q}_j$, $j = 1, \ldots, \nu$, certain discrete measures having finite support. The constants depend on the bounds of the uniformly bounded classes $\mathcal{F}^j_{\mathcal{U}}$ and $\mathcal{G}^j_{\mathcal{U}}$, $j = 1, \ldots, \nu$. Since the latter classes satisfy the condition (53) (see Examples 51 and 52), the estimate (59) implies that $\mathcal{F}_{\mathcal{U}}$ satisfies (53), too. Hence, we obtain the same estimates for mixed-integer two-stage models as in Example 52 for two-stage models without integrality requirements and in Example 51 for linear chance constrained models.

Example 54 (newsboy continued)
According to Example 15, the class $\mathcal{F}_{\mathcal{U}}$ is of the form $\mathcal{F}_{\mathcal{U}} = \{F_0(x,\cdot) = (r-c)x + c\max\{0, x-\cdot\} : x \in X \cap \mathrm{cl}\,\mathcal{U}\}$ with envelope $F_{\mathcal{F}_{\mathcal{U}}}(\xi) = r\sup_{X\cap\mathrm{cl}\,\mathcal{U}}|x| + c|\xi|$ and a uniform Lipschitz constant c. Hence, $\mathcal{F}_{\mathcal{U}}$ is a subset of $L_1(\Xi, P)$ if $\int_\Xi |\xi| dP(\xi) = \sum_{k\in\mathbb{N}} \pi_k k < \infty$. As in Example 52 we obtain

$$N(\varepsilon c, \mathcal{F}_{\mathcal{U}}, L_2(\Xi, Q)) \leq N(\varepsilon, X \cap \mathrm{cl}\,\mathcal{U}, \mathbb{R}^m) \leq C\varepsilon^{-m}$$

for each finitely discrete measure $Q \in \mathcal{P}(\Xi)$ and, hence, Theorem 50 provides the rate of convergence of the solution sets $X^*_{\mathcal{U}}(P_n(\cdot))$ of (4) with linear Ψ_P.

4.2 Scenario Generation and Reduction

Most of the numerical solution approaches for stochastic programs resort to discrete approximations of the underlying probability measure P. Several approaches have been developed for the generation or construction of discrete approximations and are in use for solving applied stochastic programming models (see the overview by Dupačová et al. (2000) and the references therein). The quantitative stability results of Section 2.3 suggest another approach, namely, to construct approximations for the original measure P such that they are close to P with respect to the corresponding probability (pseudo) metric. Let $\mathcal{F}$ be a set of measurable functions from Ξ to $\mathbb{R}$ such that the stochastic programming model (1) is stable in the sense of the Theorems 5 and 9 with respect to the (pseudo) metric

$$d_{\mathcal{F}}(P, Q) = \sup_{F\in\mathcal{F}} \Big| \int_\Xi F(\xi) d(P-Q)(\xi) \Big|$$

or some other distance bounding $d_{\mathcal{F}}(P, Q)$ from above. This means that the optimal values and the solution sets of (1) behave continuously at P when perturbing P with respect to $d_{\mathcal{F}}$.
Then it is a natural requirement to construct approximate probability distributions such that they are best approximations to P in the sense of $d_{\mathcal{F}}$. For instance, the principle of *optimal scenario generation* with a prescribed number of scenarios may be formulated as follows:
Given $P \in \mathcal{P}(\Xi)$ and $M \in \mathbb{N}$, determine a discrete probability measure $Q^* \in \mathcal{P}(\Xi)$ having M scenarios such that

$$d_{\mathcal{F}}(P, Q^*) = \min\Big\{ d_{\mathcal{F}}\Big(P, \sum_{j=1}^M q_j \delta_{\xi_j}\Big) : \sum_{j=1}^M q_j = 1, q_j \geq 0, \xi_j \in \Xi, j = 1, \ldots, M \Big\}.$$

Further constraints could be incorporated into the minimization problem, e.g., constraints implying that the scenarios exhibit a tree structure. Unfortunately, it seems to be hopeless to solve this problem for general measures P, function classes $\mathcal{F}$, supports Ξ, and large numbers M of scenarios. However, it is of course a challenging problem to develop approaches for solving the best approximation problem for more specific situations, like e.g. for the unconstrained case $\Xi = \mathbb{R}^s$, discrete measures P (involving very many scenarios) and function classes that are relevant in Section 3. An approach for solving the best approximation problem in case of $\Xi = \mathbb{R}^s$ and $\mathcal{F} = \mathcal{F}_1(\mathbb{R}^s)$ is developed in Pflug (2001).
Another important problem consists in reducing a given discrete probability measure $P = \sum_{i=1}^{N} p_i \delta_{\xi_i}$ with a (very) large number N of scenarios to a measure containing n of the original scenarios with $n << N$. Similarly as in case of optimal scenario generation, the problem of *optimal scenario reduction* may be formulated in the form

$$\min \Big\{ d_{\mathcal{F}}\Big(\sum_{i=1}^{N} p_i \delta_{\xi_i}, \sum_{j\in J} q_j \delta_{\xi_j} \Big) : J \subset \{1,\dots,N\}, |J| = n, \sum_{j\in J} q_j = 1, q_j \geq 0 \Big\}, \quad (60)$$

i.e., as a nonlinear mixed-integer program. Since its objective function is difficult to compute for general classes $\mathcal{F}$, solution methods for (60) are a challenging task. However, in the special case that $\mathcal{F} = \mathcal{F}_p(\Xi)$, for some $p \geq 1$, the objective function of (60) turns out to be the dual optimal value of the standard network flow problem (see Rachev and Rüschendorf (1998))

$$\min \Big\{ \sum_{\substack{i=1 \\ j\in J}}^{N} c_p(\xi_i, \xi_j) \|\xi_i - \xi_j\| \eta_{ij} : \eta_{ij} \geq 0, \sum_{i=1}^{N} \eta_{ij} - \sum_{j\in J} \eta_{ij} = q_j - p_i, \forall i, j \Big\},$$

where $c_p(\xi_i, \xi_j) = \max\{1, \|\xi_i\|, \|\xi_j\|\}^{p-1}$, $i, j = 1, \dots, N$, and, hence, it is a polyhedral function of q. Furthermore, in case of $\mathcal{F} = \mathcal{F}_1(\Xi)$ problem (60) simplifies considerably.

Proposition 55 *Given $J \subset \{1,\dots,N\}$ we have*

$$\min \Big\{ d_{\mathcal{F}_1(\Xi)}\Big(\sum_{i=1}^{N} p_i \delta_{\xi_i}, \sum_{j\in J} q_j \delta_{\xi_j} \Big) : \sum_{j\in J} q_j = 1, q_j \geq 0 \Big\} = \sum_{i\notin J} p_i \min_{j\in J} \|\xi_i - \xi_j\|.$$

Moreover, the minimum is attained at $\bar{q}_j = p_j + \sum_{i\in J_j} p_i$, for each $j \in J$, where $J_j := \{i \notin J : j = j(i)\}$ and $j(i) \in \arg\min_{j\in J} \|\xi_i - \xi_j\|$ for each $i \notin J$.

The proposition provides an explicit formula for the *redistribution* of the given probabilities p_i, $i = 1, \dots, N$, to the scenarios with indices in J. For its proof we refer to Theorem 2 in Dupačová et al. (2003). Due to Proposition 55 the optimal scenario reduction problem (60) in case of $\mathcal{F} = \mathcal{F}_1(\Xi)$ takes the form: Given $P \in \mathcal{P}(\Xi)$ and $n \in \mathbb{N}$, determine a solution of

$$\min\Big\{ \sum_{i \notin J} p_i \min_{j \in J} \|\xi_i - \xi_j\| : J \subset \{1, ..., N\}, |J| = n \Big\} \tag{61}$$

and compute the optimal weights $\bar{q}$ according to the redistribution rule in Proposition 55. Notice that problem (61) means that the set $\{1, \ldots, N\}$ has to be covered by a subset J of $\{1, \ldots, N\}$ and by $\{1, \ldots, N\} \setminus J$ such that $|J| = n$ and the cover has minimal cost $\sum_{i \notin J} p_i \min_{j \in J} \|\xi_i - \xi_j\|$. Hence, problem (61) is of set-covering type and, thus, $\mathcal{NP}$-hard. However, the specific structure of the objective function allows the design of fast heuristic algorithms for its approximate solution (see Dupačová et al. (2003), Heitsch and Römisch (2003)). Depending on the size of the number n of remaining scenarios, the two basic ideas are *backward reduction* and *forward selection*, respectively. In the backward reduction heuristic an index set $J = \{l_1, \ldots, l_n\}$ is determined such that

$$l_i \in \arg \min_{l \in J_r^{[i-1]}} \sum_{k \in J_r^{[i-1]} \setminus \{l\}} p_k \min_{j \notin J_r^{[i-1]} \setminus \{l\}} \|\xi_k - \xi_j\| \quad (i = 1, \ldots, n),$$

where $J_r^{[0]} = \{1, \ldots, N\}$, $J_r^{[i]} = J_r^{[i-1]} \setminus \{l_i\}$, $i = 1, \ldots, n$. In the forward selection heuristic the index set $J = \{l_1, \ldots, l_n\}$ is chosen by an opposite strategy such that

$$l_i \in \arg \min_{l \notin J_s^{[i-1]}} \sum_{k \notin J_s^{[i-1]} \cup \{l\}} p_k \min_{j \in J_s^{[i-1]} \cup \{l\}} \|\xi_k - \xi_j\| \quad (i = 1, \ldots, n)$$

holds, where $J_s^{[0]} = \emptyset$, $J_s^{[i]} = J_s^{[i-1]} \cup \{l_i\}$, $i = 1, \ldots, n$. We refer to Heitsch and Römisch (2003) for a discussion of the complexity of both heuristics, for implementation issues and encouraging numerical results.

5 Bibliographical Notes

The beginnings of approximation and estimation results in stochastic programming date back to the 1970-ies and the papers by Kall (1974) (see also the monograph Kall (1976)), Marti (1975, 1979) and Olsen (1976) on approximations, and the work of Kaňková (1977) and Wets (1979) on empirical estimation in stochastic programming. Surveys on stability were published by Dupačová (1990) and Schultz (2000). The notion of stability of stochastic

programs appeared first in Bereanu (1975), in the context of the distribution problem, and in Kaňková (1978), where stability of minima of more general stochastic programming models was studied with respect to weak convergence of measures for the first time.
Later Dupačová (1984, 1987) and Wang (1985) studied the stability of stochastic programs with respect to changes of finite-dimensional parameters in the underlying probability distribution. Kall and Stoyan (1982), Salinetti (1983) and Römisch (1981, 1985) dealt with discrete approximations to stochastic programs. Further early work has been done in the surveys by Wets (1983, 1989) and in Friedrich and Tammer (1981) (on stability), Birge and Wets (1986) (on discrete approximation schemes), Römisch (1986b), Kall (1987), Robinson and Wets (1987), Römisch and Wakolbinger (1987), Vogel (1988), Kall, Ruszczyński and Frauendorfer (1988) (on discrete approximations), Dupačová and Wets (1988), Shapiro (1989) and Wang (1989). The landmark papers by Kall (1987) and by Robinson and Wets (1987) address qualitative stability results for optimal values and solution sets with respect to weak convergence of measures. This line of research was continued in the important work by Artstein and Wets (1994) and in Vogel (1992), Schultz (1992, 1995), Lucchetti and Wets (1993), Wang (1995), Wets (1998), Zervos (1999) and Riis and Schultz (2002). Attempts to quantify such stability results using distances of probability measures were started in Römisch (1986b), Römisch and Wakolbinger (1987) and continued in Römisch and Schultz (1991a–c, 1993, 1996), Artstein (1994), Kaňková (1994b, 1998), Shapiro (1994), Fiedler and Römisch (1995), Schultz (1996), Henrion and Römisch (1999, 2000), Dentcheva (2000) and Rachev and Römisch (2002).
Most of the stability studies allow for general perturbations of the underlying probability measure and develop a general framework for both *discrete* and *statistical* approximations of stochastic programs. Nevertheless, these two kinds of approximations developed independently by exploiting their specific structures (e.g. bounding techniques on the one hand and asymptotic statistical arguments on the other hand). For (discrete) approximations we mention the work in Birge and Wets (1986), Kall, Ruszczyński and Frauendorfer (1988), Lepp (1990), Birge and Qi (1995a–b), Frauendorfer (1992, 1996), Kall (1998). In parallel, statistical inference in stochastic programming models was studied intensively. After the early work by Kaňková and Wets, many authors contributed to this line of research on asymptotic properties of statistical estimators, e.g., their consistency, rates of convergence and limit theorems. We mention, in particular, the work of Dupačová and Wets (1988), Vogel (1988), Shapiro (1989, 1990, 1991, 1996, 2000), King (1989), Kaňková (1990, 1994), King and Wets (1991), Wets (1991), Ermoliev and Norkin (1991), Norkin (1992), King and Rockafellar (1993), Rubinstein and Shapiro (1993), Bouza (1994), Geyer (1994), Artstein and Wets (1995), Kaniovski, King and Wets (1995), Lachout (1995), Pflug (1995, 1999), Robinson (1996), Gröwe (1997), Pflug, Ruszczyński and Schultz (1998a,b), Mak, Morton and Wood (1999), Shapiro and Homem-de-Mello (2000).

Another line of research on approximations of stochastic programs is based on the convergence (almost surely, in probability and in distribution) of measurable set-valued mappings and on the epi-convergence of integrands. Here, we mention the fundamental paper by Salinetti and Wets (1986) and the work of Salinetti (1981, 1983), Römisch (1986a), Vogel (1988, 1992, 1994, 1995), Wets (1991), Hess (1996) and the recent papers by Korf and Wets (2000, 2001) and by Vogel and Lachout (2000).
Much is known on the stability of values and solutions of classical two-stage stochastic programs (Section 3.1). The situation is already different for the stability of solutions to chance constrained models and even more to mixed-integer two-stage models. The stability of multi-stage stochastic programs is widely open, especially in the mixed-integer case. Another open matter are the stability effects of incorporating risk functionals into stochastic programming models (cf. Section 2.4).
The paper by Rachev and Römisch (2002) provides an important source for the material presented in this chapter, in particular, for the Sections 2.2, 2.3, 4.1 and parts of the Sections 3.1 and 3.2. Some of the results are directly taken from that paper, namely, Theorems 5, 9, 23 and 39. Some other results represent modified or extended versions of those in Rachev and Römisch (2002) (e.g. Theorems 35 and 50). Theorems 13 and 24 are due to work in preparation by Römisch and Wets. Corollary 45 and Theorem 47 are taken from Henrion and Römisch (1999) and the Corollaries 42 and 44 from Römisch and Schultz (1991c). The Example 41 is due to Henrion and the notion of a Lipschitz continuous risk functional goes back to Pflug (2002).

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft, in particular, by the Schwerpunktprogramm *Online Optimization of Large Scale Systems.* The author wishes to thank Darinka Dentcheva (Stevens Institute of Technology) and René Henrion (Weierstrass Institute Berlin) for invaluable discussions, in particular, on Steiner selections of set-valued maps and on the stability of chance constrained models, respectively. Further thanks are due to Georg Pflug (University of Vienna) and Roger Wets (University of California at Davis) for valuable conversations on empirical processes and risk functionals and on the stability of stochastic programs, respectively.

References

Artstein, Z. (1994). Sensitivity with respect to the underlying information in stochastic programs, *Journal of Computational and Applied Mathematics* 56, 127–136.

Artstein, Z.; Wets, R.J-B (1994). Stability results for stochastic programs and sensors, allowing for discontiuous objective functions, *SIAM Journal on Optimization* 4, 537–550.

Artstein, Z.; Wets, R.J-B (1995). Consistency of minimizers and the SLLN for stochastic programs, *Journal of Convex Analysis* 2, 1–17.

Artzner, P.; Delbaen, F.; Eber, J.-M.; Heath, D. (1999). Coherent measures of risk, *Mathematical Finance* 9, 203–228.

Attouch, H.; Wets, R.J-B (1993). Quantitative stability of variational systems II. A framework for nonlinear conditioning, *SIAM Journal on Optimization* 3, 359–381.

Bank, B.; Guddat, J.; Klatte, D.; Kummer, B.; Tammer, K. (1982). *Non-Linear Parametric Optimization*, Akademie-Verlag, Berlin.

Bereanu, B. (1975). Stable stochastic linear programs and applications, *Mathematische Operationsforschung und Statistik* 6, 593–607.

Berge, C. (1963). Topological spaces, Oliver & Boyd, Edinburgh-London.

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

Billingsley, P.; Topsøe, F. (1967). Uniformity in weak convergence, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 7, 1–16.

Birge, J.R.; Qi, L. (1995a). Subdifferential convergence in stochastic programs, *SIAM Journal on Optimization* 5, 436–453.

Birge, J.R.; Qi, L. (1995b). Continuous approximation schemes for stochastic programs, *Annals of Operations Research* 56, 15–38.

Birge, J.R.; Wets, R.J-B (1986). Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse, *Mathematical Programming Study* 27, 54–102.

Blair, C.E.; Jeroslow, R.G. (1977). The value function of a mixed integer program, *Discrete Mathematics* 19, 121–138.

Bonnans, J.F.; Shapiro, A. (2000). *Perturbation Analysis of Optimization Problems*, Springer-Verlag, New York.

Bouza, C. (1994). Approximation of the value of the optimal solution in stochastic programming, *Revista Investigacion Operacional* 15, 13–33.

Delbaen, F. (2002). Coherent risk measures on general probability spaces, in: *Advances in Finance and Stochastics* (K. Sandmann and P.J. Schönbucher Eds.), Springer, Berlin, 1–37.

Dentcheva, D. (2000). Regular Castaing representations of multifunctions with applications to stochastic programming, *SIAM Journal on Optimization* 10, 732–749.

Dentcheva, D.; Römisch, W. (2000). Differential stability of two-stage stochastic programs, *SIAM Journal on Optimization* 11, 87–112.

Dudley, R.M. (1984). A course on empirical processes, in: École d'Été de Probabilities de Saint-Flour XII – 1982, *Lecture Notes in Mathematics* 1097, Springer, Berlin, 2–142.

Dudley, R.M. (1989). *Real Analysis and Probability*, Wadsworth & Brooks/Cole, Pacific Grove.

Dupačová, J. (1984). Stability in stochastic programming with recourse - estimated parameters, *Mathematical Programming* 28, 72–83.

Dupačová, J. (1987). Stochastic programming with incomplete information: A survey of results on postoptimization and sensitivity analysis, *Optimization* 18, 507–532.

Dupačová, J. (1990). Stability and sensitivity analysis for stochastic programming, *Annals of Operations Research* 27, 115–142.

Dupačová, J. (1994). Applications of stochastic programming under incomplete information, *Journal of Computational and Applied Mathematics* 56, 113–125.

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

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

Dupačová, J.; Wets, R.J-B (1988). Asymptotic behaviour of statistical estimators and of optimal solutions of stochastic optimization problems, *The Annals of Statistics* 16, 1517–1549.

Ermoliev, Y.M.; Norkin, V.I. (1991). Normalized convergence in stochastic optimization, *Annals of Operations Research* 30, 187–198.

Fiedler, O.; Römisch, W. (1995). Stability in multistage stochastic programming, *Annals of Operations Research* 56, 79–93.

Föllmer, H.; Schied, A. (2002). Convex risk measures and trading constraints, *Finance and Stochastics* 6, 429–447.

Fortet, R.; Mourier, E. (1953). Convergence de la répartition empirique vers la répartition théorique, *Ann. Sci. Ecole Norm. Sup.* 70, 266–285.

Frauendorfer, K. (1992). *Stochastic Two-Stage Programming*, Lecture Notes in Economics and Mathematical Systems Vol. 392, Springer, Berlin.

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

Friedrich, B.; Tammer, K. (1981). Untersuchungen zur Stabilität einiger Ersatzprobleme der stochastischen Optimierung in bezug auf Änderungen des zugrunde gelegten Zufallsvektors, *Wissenschaftliche Zeitschrift der Humboldt-Universität Berlin*, Math.-Naturwiss. Reihe 30, 373–376.

Geyer, C.J. (1994). On the asymptotics of constrained M-estimation, *The Annals of Statistics* 22, 1993–2010.

Gröwe, N. (1997). Estimated stochastic programs with chance constraints, *European Journal of Operational Research* 101, 285–305.

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

Henrion, R. (2000). Qualitative stability of convex programs with probabilistic constraints, in: *Optimization* (V.H. Nguyen, J.-J. Strodiot and P. Tossings Eds.), Lecture Notes in Economics and Mathematical Systems Vol. 481, Springer, Berlin, 164–180.

Henrion, R.; Römisch, W. (1999). Metric regularity and quantitative stability in stochastic programs with probabilistic constraints, *Mathematical Programming* 84, 55–88.

Henrion, R.; Römisch, W. (2000). Stability of solutions to chance constrained stochastic programs, in: *Parametric Optimization and Related Topics V* (J. Guddat, R. Hirabayashi, H.Th. Jongen and F. Twilt Eds.), Peter Lang Verlag, Frankfurt a.M., 95–114.

Henrion, R.; Römisch, W. (2002). Hölder and Lipschitz stability of solution sets in programs with probabilistic constraints, Preprint No. 798, Weierstrass Institute for Applied Analysis and Stochastics Berlin, and submitted.

Hess, C. (1996). Epi-convergence of sequences of normal integrands and strong consistency of the maximum likelihood estimator, *The Annals of Statistics* 24, 1298–1315.

Kall, P. (1974). Approximations to stochastic programs with complete recourse, *Numerische Mathematik* 22, 333–339.

Kall, P. (1976). *Stochastic Linear Programming*, Springer, Berlin.

Kall, P. (1987). On approximations and stability in stochastic programming, in: *Parametric Optimization and Related Topics* (J. Guddat, H.Th. Jongen, B. Kummer and F. Nožička, Eds.), Akademie-Verlag, Berlin, 387–407.

Kall, P.; Ruszczyński, A.; Frauendorfer, K. (1988). Approximation techniques in stochastic programming, in: *Numerical Techniques for Stochastic Optimization* (Y. Ermoliev and R.J-B Wets Eds.), Springer, Berlin, 33–64.

Kall, P. (1998). Bounds for and approximations to stochastic linear programs with recourse, in : *Stochastic Programming Methods and Technical Applications* (K. Marti and P. Kall Eds.), Lecture Notes in Economics and Mathematical Systems Vol. 458, Springer, Berlin, 1–21.

Kaniovski, Y.M.; King, A.J.; Wets, R.J-B (1995). Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems, *Annals of Operations Research* 56, 189–208.

Kaňková, V. (1977). Optimum solution of a stochastic optimization problem with unknown parameters, in: Transactions of the 7th Prague Conference on *Information Theory, Statistical Decision Functions and Random Processes*, Academia, Prague, 239–244.

Kaňková, V. (1978). Stability in the stochastic programming, *Kybernetika* 14, 339–349.

Kaňková, V. (1990). On the convergence rate of empirical estimates in chance constrained stochastic programming, *Kybernetika* 26, 449–461.

Kaňková, V. (1994). A note on estimates in stochastic programming, *Journal of Computational and Applied Mathematics* 56, 97–112.

Kaňková, V. (1994). On distribution sensitivity in stochastic programming, Research Report No. 1826, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic.

Kaňková, V. (1998). A note on multifunctions in stochastic programming, in: *Stochastic Programming Methods and Technical Applications* (K. Marti and P. Kall Eds.), Lecture Notes in Economics and Mathematical Systems Vol. 458, Springer, Berlin, 154–168.

King, A.J. (1989). Generalized delta theorems for multivalued mappings and measurable selections, *Mathematics of Operations Research* 14, 720–736.

King, A.J.; Rockafellar, R.T. (1993). Asymptotic theory for solutions in statistical estimation and stochastic programming, *Mathematics of Operations Research* 18, 148–162.

King, A.J.; Wets, R.J-B (1991). Epi-consistency of convex stochastic programs, *Stochastics and Stochastics Reports* 34, 83–92.

Klatte, D. (1987). A note on quantitative stability results in nonlinear optimization, in: *Proceedings 19. Jahrestagung Mathematische Optimierung* (K. Lommatzsch Ed.), Humboldt-Universität Berlin, Sektion Mathematik, Seminarbericht Nr. 90, 77–86.

Klatte, D. (1994). On quantitative stability for non-isolated minima, *Control and Cybernetics* 23, 183–200.

Korf, L.; Wets, R.J-B (2000). Random lsc functions: scalarization, *Stochastic Programming E-Print Series* 7.

Korf, L.; Wets, R.J-B (2001). Random lsc functions: An ergodic theorem, *Mathematics of Operations Research* 26, 421–445.

Lachout, P. (1995). On multifunction transforms of probability measures, *Annals of Operations Research* 56, 241–249.

Lepp, R. (1990). Approximations to stochastic programs with complete recourse, *SIAM Journal on Control and Optimization* 28, 382–394.

Louveaux, F.; Schultz, R. (2003). *Stochastic Integer Programming*, Chapter 4, this volume.

Lucchetti, R.; Wets, R.J-B (1993). Convergence of minima of integral functionals, with applications to optimal control and stochastic optimization, *Statistics & Decisions* 11, 69–84.

Lucchetti, R.; Salinetti, G.; Wets, R.J-B (1994). Uniform convergence of probability measures: Topological criteria, *Journal of Multivariate Analysis* 51, 252–264.

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

Marti, K. (1975). Approximationen der Entscheidungsprobleme mit linearer Ergebnisfunktion und positiv homogener, subadditiver Verlustfunktion, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 31, 203–233.

Marti, K. (1979). *Approximationen stochastischer Optimierungsprobleme*, Mathematical Systems in Economics Vol. 43, A. Hain, Königstein/Ts.

Mordukhovich, B.S. (1994a). Generalized differential calculus for nonsmooth and set-valued mappings, *Journal of Mathematical Analysis and Applications* 183, 250–288.

Mordukhovich, B.S. (1994b). Lipschitzian stability of constraint systems and generalized equations, *Nonlinear Analysis, Theory, Methods & Applications* 22, 173–206.

Norkin, V.I. (1992). Convergence of the empirical mean method in statistics and stochastic programming, *Cybernetics and Systems Analysis* 28, 253–264.

Ogryczak, W.; Ruszczyński, A. (1999). From stochastic dominance to mean-risk models: semideviation as risk measures, *European Journal of Operational Research* 116, 33–50.

Olsen, P. (1976). Discretizations of multistage stochastic programming problems, *Mathematical Programming Study* 6, 111–124.

Pflug, G. (1995). Asymptotic stochastic programs, *Mathematics of Operations Research* 20, 769–789.

Pflug, G. (1999). Stochastic programs and statistical data, *Annals of Operations Research* 85, 59–78.

Pflug, G. (2001). Scenario tree generation for multiperiod financial optimization by optimal discretization, *Mathematical Programming*, Series B, 89, 251–271.

Pflug, G. (2003). *Stochastic Optimization and Statistical Inference*, Chapter 7, this volume.

Pflug, G.; Ruszczyński, A.; Schultz, R. (1998a). On the Glivenko-Cantelli problem in stochastic programming: Linear recourse and extensions, *Mathematics of Operations Research* 23, 204–220.

Pflug, G.; Ruszczyński, A.; Schultz, R. (1998b). On the Glivenko-Cantelli problem in stochastic programming: Mixed-integer linear recourse, *Mathematical Methods of Operations Research* 47, 39–49.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*, NSF-CBMS Regional Conference Series in Probability and Statistics Vol. 2, Institute of Mathematical Statistics.

Prékopa, A. (1995). *Stochastic Programming*, Kluwer, Dordrecht.

Rachev, S.T. (1991). *Probability Metrics and the Stability of Stochastic Models*, Wiley, Chichester.

Rachev, S.T.; Römisch, W. (2002). Quantitative stability in stochastic programming: The method of probability metrics, *Mathematics of Operations Research* 27, 792–818.

Rachev, S.T.; Rüschendorf, L. (1998). Mass Transportation Problems, Vol. I: Theory, Springer, New York.

Riis, M.; Schultz, R. (2002). Applying the minimum risk criterion in stochastic recourse programs, *Computational Optimization and Applications* (to appear).

Robinson, S.M. (1987). Local epi-continuity and local optimization, *Mathematical Programming* 37, 208–223.

Robinson, S.M. (1996). Analysis of sample-path optimization, *Mathematics of Operations Research* 21, 513–528.

Robinson, S.M.; Wets, R.J-B (1987). Stability in two-stage stochastic programming, *SIAM Journal on Control and Optimization* 25, 1409–1416.

Rockafellar, R.T.; Wets, R.J-B (1998). *Variational Analysis*, Springer, Berlin.

Rockafellar, R.T.; Uryasev, S. (2002). Conditional value-at-risk for general loss distributions, *Journal of Banking & Finance* 26, 1443–1471.

Römisch, W. (1981). On discrete approximations in stochastic programming, in: Proceedings 13. Jahrestagung *Mathematische Optimierung* (K. Lommatzsch Ed.), Humboldt-Universität Berlin, Sektion Mathematik, Seminarbericht Nr. 39, 166–175.

Römisch, W. (1985). An approximation method in stochastic optimization and control, in: *Mathematical Control Theory*, Banach Center Publications Vol. 14, PWN, Warsaw, 477–490.

Römisch, W. (1986a). On the convergence of measurable selections and an application to approximations in stochastic optimization, *Zeitschrift für Analysis und Anwendungen* 5, 277–288.

Römisch, W. (1986b). On convergence rates of approximations in stochastic programming, in: Proceedings 17. Jahrestagung *Mathematische Optimierung* (K. Lommatzsch Ed.), Humboldt-Universität Berlin, Sektion Mathematik, Seminarbericht Nr. 80, 82–91.

Römisch, W.; Schultz, R. (1991a). Distribution sensitivity in stochastic programming, *Mathematical Programming* 50, 197–226.

Römisch, W.; Schultz, R. (1991b). Distribution sensitivity for certain classes of chance-constrained models with application to power dispatch, *Journal of Optimization Theory and Applications* 71, 569–588.

Römisch, W.; Schultz, R. (1991c). Stability analysis for stochastic programs, *Annals of Operations Research* 30, 241–266.

Römisch, W.; Schultz, R. (1993), Stability of solutions for stochastic programs with complete recourse, *Mathematics of Operations Research* 18, 590–609.

Römisch, W.; Schultz, R. (1996). Lipschitz stability for stochastic programs with complete recourse, *SIAM Journal on Optimization* 6, 531–547.

Römisch, W.; Wakolbinger, A. (1987). Obtaining convergence rates for approximations in stochastic programming, in: *Parametric Optimization and Related Topics* (J. Guddat, H.Th. Jongen, B. Kummer and F. Nožička Eds.), Akademie-Verlag, Berlin, 327–343.

Rubinstein, R.Y.; Shapiro, A. (1993). *Discrete Event Systems, Sensitivity Analysis and Stochastic Optimization by the Score Function Method*, Wiley, Chichester.

Ruszczyński, A.; Shapiro, A. (2003). *Stochastic Programming Models*, Chapter 1, this volume.

Salinetti, G. (1981). Convergence of measurable multifunctions: An application to stochastic optimization problems, in: *Mathematical Programming and its Economic Applications* (G. Castellani and P. Mazzoleni Eds.), Angeli, Milano, 757–774.

Salinetti, G. (1983). Approximations for chance-constrained programming problems, *Stochastics* 10, 157–179.

Salinetti, G.; Wets, R.J-B (1986). On the convergence in distribution of measurable multifunctions (random sets), normal integrands, stochastic processes and stochastic infima, *Mathematics of Operations Research* 11, 385–419.

Schultz, R. (1992). Continuity and stability in two-stage stochastic integer programming, in: *Stochastic Optimization, Numerical Methods and Technical Applications* (K. Marti Ed.), Lecture Notes in Economics and Mathematical Systems Vol. 379, Springer, Berlin, 81–92.

Schultz, R. (1994). Strong convexity in stochastic programs with complete recourse, *Journal of Computational and Applied Mathematics* 56, 3–22.

Schultz, R. (1995). On structure and stability in stochastic programs with random technology matrix and complete integer recourse, *Mathematical Programming* 70, 73–89.

Schultz, R. (1996). Rates of convergence in stochastic programs with complete integer recourse *SIAM Journal on Optimization* 6, 1138–1152.

Schultz, R. (2000). Some aspects of stability in stochastic programming, *Annals of Operations Research* 100, 55–84.

Shapiro, A. (1989). Asymptotic properties of statistical estimators in stochastic programming, *The Annals of Statistics* 17, 841–858.

Shapiro, A. (1990). On differential stability in stochastic programming, *Mathematical Programming* 47, 107–116.

Shapiro, A. (1991). Asymptotic analysis of stochastic programs, *Annals of Operations Research* 30, 169–186.

Shapiro, A. (1994). Quantitative stability in stochastic programming, *Mathematical Programming* 67, 99–108.

Shapiro, A. (1996). Simulation-based optimization – Concergence analysis and statistical inference, *Communications in Statistics – Stochastic Models* 12, 425–454.

Shapiro, A. (2000). Statistical inference of stochastic optimization problems, in: *Probabilistic Constrained Optimization* (S.P. Uryasev Ed.), Kluwer, Dordrecht, 282–307.

Shapiro, A. (2003). *Monte Carlo Sampling Methods*, Chapter 6, this volume.

Shapiro, A.; Homem-de-Mello, T. (2000). On rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs, *SIAM Journal on Optimization* 11, 70–86.

Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes, *Annals of Probability* 22, 28–76.

Talagrand, M. (1996). The Glivenko-Cantelli problem, ten years later, *Journal of Theoretical Probability* 9, 371–384.

Topsøe, F. (1967). On the connection between P-continuity and P-uniformity in weak convergence, *Probability Theory and its Applications* 12, 281–290.

Topsøe, F. (1977). Uniformity in convergence of measures, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 39, 1–30.

van der Vaart, A.W. (1998). *Asymptotic Statistics*, Cambridge University Press.

van der Vaart, A.W.; Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*, Springer, New York.

Vogel, S. (1988). Stability results for stochastic programming problems, *Optimization* 19, 269–288.

Vogel, S. (1992). On stability in multiobjective programming – A stochastic approach, *Mathematical Programming* 56, 91–119.

Vogel, S. (1994). A stochastic approach to stability in stochastic programming, *Journal of Computational and Applied Mathematics* 56, 65–96.

Vogel, S. (1995). On stability in stochastic programming – Sufficient conditions for continuous convergence and epi-convergence, Preprint No. M 12/95, Technical University Ilmenau, Faculty of Mathematics and Natural Sciences.

Vogel, S.; Lachout, P. (2000). On continuous convergence and epi-convergence of random functions – I and II, Research Report No. 1988 and 1989, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic.

Walkup, D.; Wets, R.J-B (1969a). Lifting projections of convex polyhedra, *Pacific Journal of Mathematics* 28, 465–475.

Walkup, D.; Wets, R.J-B (1969b). A Lipschitzian characterization of convex polyhedra, *Proceedings of the American Mathematical Society* 23, 167–173.

Wang, J. (1985). Distribution sensitivity analysis for stochastic programs with complete recourse, *Mathematical Programming* 31, 286–297.

Wang, J. (1989). Continuity of feasible solution sets of probabilistic constrained programs, *Journal of Optimization Theory and Applications* 63, 79–89.

Wang, J. (1995). Stability of multistage stochastic programming, *Annals of Operations Research* 56, 313–322.

Wets, R.J-B (1974). Stochastic programs with fixed recourse: The equivalent deterministic program, *SIAM Review* 16, 309–339.

Wets, R.J-B (1979). A statistical approach to the solution of stochastic programs with (convex) simple recourse, Working Paper, Department of Mathematics, University of Kentucky, Lexington.

Wets, R.J-B (1983). Stochastic programming: Solution techniques and approximation schemes, in: *Mathematical Programming: the State-of-the-Art* (A. Bachem, M. Grötschel and B. Korte Eds.), Springer, Berlin, 566–603.

Wets, R.J-B (1989). Stochastic programming, in: *Handbooks in Operations Research and Management Science, Vol. 1, Optimization* (G.L. Nemhauser, A.H.G. Rinnooy Kan, M.J. Todd Eds.), North-Holland, Amsterdam, 573–629.

Wets, R.J-B (1991). Constrained estimation: Consistency and asymptotics, *Applied Stochastic Models and Data Analysis* 7, 17–32.

Wets, R.J-B (1998). Stochastic programs with chance constraints: Generalized convexity and approximation issues, in: *Generalized Convexity, Generalized Monotonicity: Recent Results* (J.-P. Crouzeix, J.-E. Martínez-Legaz, M. Volle Eds.), Kluwer, Dordrecht, 61–74.

Zervos, M. (1999). On the epiconvergence of stochastic optimization problems, *Mathematics of Operations Research* 24, 495–508.

Zolotarev, V.M. (1983). Probability metrics, *Theory of Probability and its Applications* 28, 278–302.