Instituto de Matemática Pura e Aplicada

Doctoral Thesis

# SOME ISSUES ON STOCHASTIC PROGRAMMING PROBLEMS

Marcus de Mendes Caldas Raymundo Reaiche

Rio de Janeiro
2017

Instituto de Matemática Pura e Aplicada

Marcus de Mendes Caldas Raymundo Reaiche

# SOME ISSUES ON STOCHASTIC PROGRAMMING PROBLEMS

Thesis presented to the Post-graduate Program in Mathematics at Instituto de Matemática Pura e Aplicada as partial fulfillment of the requirements for the degree of Doctor of Science in Mathematics.

**Advisor:** Alfredo Noel Iusem
**Co-advisor:** Alexander Shapiro

**Rio de Janeiro**
**August 16, 2017**

# Agradecimentos

Agradeço ao meu orientador, Alfredo Noel Iusem, pela atenciosa orientação durante os meus anos de doutorado. Alfredo sempre esteve disponível para conversarmos sobre a pesquisa. Ele teve um papel decisivo quando percebeu nos meados do programa que eu estava tentando resolver um problema sem muito sucesso. Seus incentivos constantes e conselhos preciosos fizeram com que essa jornada tivesse um final feliz. A ele o meu muito obrigado!

I would also like to thank my co-advisor Alexander Shapiro from Georgia Institute of Technology, USA. Both problems that were solved in this thesis were proposed by him. Shapiro gave me a lot of freedom and flexibility to explore these problems. I am particularly grateful for his guidance, patience and kindness while I was working with him for 1 year in Atlanta.

Agradeço aos membros da banca examinadora deste trabalho pelos seus comentários e sugestões que contribuíram para o seu aprimoramento.

Agradeço aos professores do IMPA por tudo o que me ensinaram durante todos esses anos desde os meus tempos de mestrado. Agradeço em especial aos professores Roberto Imbuzeiro, Claudia Sagastizábal e Alfredo Iusem por terem investido tanto tempo em mim. Fiz inúmeros cursos com o Roberto o que foi extremamente útil para a obtenção de alguns resultados da tese.

Gostaria também de agradecer a todos os meus amigos do IMPA e da Georgia Tech com quem tive o prazer de compartilhar inúmeros momentos de reflexão e descontração durante os anos de doutorado. Agradeço, em particular, ao meu amigo Philip Thompson por todas as conversas que tivemos ao longo desses anos tanto sobre otimização quanto sobre outros assuntos mais mundanos.

Gostaria também de agradecer ao meu amigo Asteróide Santanna e a sua esposa Denise Batista que tivemos, minha esposa e eu, o imenso prazer de conhecer durante a nossa estadia em Atlanta. Como não registrar aqui as vindas em Fev/2016 da amada Sophia, filha do Aster e da Denise, e, logo em seguida, em Abr/2016 do nosso amado filho Bruno. Quantos aprendizados, sustos e alegrias nós compartilhamos juntos!

Agradeço também aos meus colegas e amigos do BNDES e ao apoio recebido pelo Banco para a minha formação. Gostaria de agradecer em especial aos amigos Gil Bernardo, Fábio Giambiagi, Lavínia Castro, Felipe Vilhena, Felipe Pinheiro e Diogo Gobira. Cada um teve um papel importante para a realização deste trabalho.

Agradeço ao CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico – pelo apoio dado para a realização do Doutorado Sanduíche no exterior.

Agradeço também aos meus pais pelo apoio e incentivos constantes dados durante toda a minha vida. Muito obrigado por sempre terem acreditado em mim e por me mostrarem as coisas pelas quais realmente vale a pena lutar nessa na vida. Voces são os meus exemplos de vida!

Finalmente, agradeço a minha esposa, Suzana Alves, meu amor, minha companheira de vida, que esteve sempre ao meu lado; celebrando junto nos bons momentos e me apoiando e consolando naqueles mais difíceis. Obrigado por tornar a minha vida tão maravilhosa!

# Abstract

In this thesis we study *sample complexity* issues of a *Monte Carlo* sampling-based approach that is used for approximating general stochastic programming problems. The thesis is divided into two parts.

In the first part of the text, we derive sample complexity estimates for a class of *risk averse* stochastic programming problems assuming standard regularity conditions. We consider the class of *Optimized Certainty Equivalent* (OCE) risk measures. We derive estimates either for *static* or *two-stage* problems as for *dynamic* or $T$*-stage* problems ($T \in \mathbb{N}$). Our results extend the ones obtained previously for risk neutral stochastic programming problems in the static and dynamic settings. In particular, we derive *exponential rates* of convergence of the *statistical estimators* associated with the approximate problem to their true counterparts. We note that the constants associated with the exponential rate of convergence deteriorate depending of how large is the *Lipschitz constant* of the problem's risk measure and of the number of stages $T$. In this case, our results indicate that in the risk averse setting one needs to construct a *scenario tree* using a relatively larger number of samples than in the risk neutral setting in order to obtain a good approximation for the solution of the original problem.

In the second part of the thesis, we derive a tight *lower bound* for the sample complexity of a class of risk neutral $T$-stage stochastic programming problems. An upper bound estimate was derived previously in the literature. Treating the number of stages $T$ as a varying parameter, our result shows that the number of scenarios needed for approximating the true problem with a desirable level of confidence *grows* more rapidly with respect to $T$ than *exponentially*. This work was published in [53].

**Keywords:** Stochastic programming, Monte Carlo sampling, Sample average method, Sample complexity, Risk averse optimization

# Resumo

Nesta tese estudamos questões de *complexidade amostral* de métodos de simulação de *Monte Carlo* que são usados a fim de aproximar problemas gerais de otimização estocástica. A tese se divide em duas partes.

Na primeira parte do texto, derivamos estimativas de *complexidade amostral* para uma classe de problemas de programação estocástica *avessos ao risco* assumindo condições de regularidade padrão. Consideramos a classe de medidas de risco conhecida como *Equivalente Certo Otimizado* (ECO). Derivamos estimativas de complexidade amostral tanto para problemas *estáticos* ou de *dois estágios* quanto para problemas *dinâmicos* ou de *multiestágios*. Nossos resultados se assemelham aos obtidos anteriormente para problemas de otimização estocástica neutros ao risco estáticos e dinâmicos. Em particular, derivamos *taxas exponenciais* de convergência dos *estimadores estatísticos* associados ao problema de aproximao às suas verdadeiras contrapartes. Observamos que as constantes associadas à taxa exponencial de convergência deterioram-se dependendo de quão grande é a *constante de Lipschitz* da medida de risco do problema e do número de estágios $T$ do problema. Neste caso, nossos resultados indicam que para problemas risco avessos deve-se construir uma *árvore de cenários* usando-se um número relativamente maior de amostras do que para problemas neutros ao risco a fim de se obter uma solução aproximada do problema original.

Na segunda parte da tese, derivamos um *limite inferior* justo para a complexidade amostral de uma classe de problemas multiestgios de programao estocstica neutros ao risco. Uma estimativa do limite superior foi derivada anteriormente na literatura. Tratando o nmero de estgios $T$ como um parâmetro variável, nosso resultado mostra que o número de cenários necessários para aproximar o problema original com um nvel de confiança especificado *cresce* mais rapidamente em relação a $T$ do que *exponencialmente*. Esse trabalho foi publicado em [53].

**Palavras-chave:** Programação estocástica, Simulação de Monte Carlo, Método da média amostral, Complexidade amostral, Otimização avessa ao risco

# Contents

Abstract . . . v
Resumo . . . vii

**1 Introduction** **1**

**2 Background material and preliminary results** **25**
2.1 Risk neutral stochastic programming problems . . . 25
2.1.1 The static case . . . 25
2.1.2 The dynamic case . . . 36
2.2 Scenario Trees . . . 62
2.3 Quantiles . . . 65
2.4 Sub-Gaussian and $\psi_2$-random variables . . . 68
2.5 Convex analysis . . . 72
2.6 Set-valued analysis . . . 89
2.6.1 Continuity of optimal value functions . . . 91
2.6.2 Measurability of multifunctions . . . 94
2.7 Risk measures . . . 98
2.8 Miscellany . . . 103

**3 Sample complexity for static problems with OCE risk measures** **107**
3.1 Optimized certainty equivalent risk measures . . . 107
3.2 Static problems with OCE risk measures . . . 125
3.3 Sample complexity results for static problems . . . 133

**4 Sample complexity for dynamic problems with OCE risk measures** **155**

**5 A lower bound for the sample complexity of a class of risk neutral dynamic problems** **173**
5.1 Introduction . . . 173
5.2 The main result . . . 179

**Bibliography** **187**

# CHAPTER 1

## Introduction

Stochastic programming problems are a class of optimization problems involving uncertainty parameters. In many real-life applications, uncertainty is present in the decision making process. At the time a decision must be made, like producing a good, relevant data may not be known by the decision maker, like future demands of the good. In stochastic programming formulations, the uncertainty quantities are modeled via the theory of probability. One usually assumes that the (joint) probability distribution of the random quantities is known or can be estimated from historical data. Stochastic programming provides useful tools that are used in real-life applications. These models occur in many areas of science and engineering, like energy [16, 26, 46, 72, 76], finance [15, 20, 49, 57] and transportation [24, 27], to mention a few.

There exists a myriad of stochastic programming paradigms for dealing with a multitude of decision making situations. The models can be classified as two-stage and multistage models with recourse, linear and nonlinear models, models with chance constraints and models with deterministic constraints, risk neutral and risk averse models, etc. We begin our exposition by considering the following general stochastic programming problem

$$\min_{x\in X} \left\{ f(x) := \mathbb{E}F(x,\xi) \right\}, \tag{1.0.1}$$

where $\xi = (\xi_1, \ldots, \xi_d)$ is a random vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$; $x \in \mathbb{R}^n$ represents the decision variables; $X \subseteq \mathbb{R}^n$ is the feasible set and $F : \mathbb{R}^n \times \mathbb{R}^d \to \overline{\mathbb{R}}$ is a measurable function. At the time the decision $x$ must be made, the optimizer does not know the value assumed by the random vector $\xi$, only its probability distribution $\mathbb{P}$. We assume that $\mathbb{P}$ does not depend on the decision $x$, the set $X$ is deterministic and the function $F$ is known or we can evaluate $F(x,\xi)$ with low computational cost[1], for each $x \in \mathbb{R}^n$ and $\xi \in \mathbb{R}^d$. Before knowing the realization of the random vector $\xi$, $F(x,\xi)$ represents a random cost, for each $x \in \mathbb{R}^n$. One way to summarize the random cost into a real number is by considering the expected cost function $f(x) := \mathbb{E}F(x,\xi)$. This is precisely the case in risk neutral problems like problem (1.0.1), where the objective function is the expected cost function. Of course, problem (1.0.1) is well-defined if and only if the expected value of $F(x,\xi)$ is well-defined[2], for each $x \in X$.

Typically the decision maker does not know, a priori, a closed form formula[3] of function $f$, and so, he must somehow evaluate it. Let us discuss some possibilities to evaluate $f$. If the random vector $\xi$ has an absolutely continuous distribution with respect to the Lebesgue measure on $\mathbb{R}^d$, then $f(x)$ is a $d$-dimensional integral, for each $x \in \mathbb{R}^n$. In very simple situations, it is possible to obtain a closed form formula for $f(x)$ using *symbolic integration*, but this is atypical for stochastic programming problems. Another approach for evaluating $f(x)$ would be via *numerical integration* techniques, like *quadrature methods*. It is well-known (see [23, Section 1.2]) that the integration problem suffers from the *dimensional effect* or *the curse of dimensionality*. Numerical integration techniques tends to deteriorate rapidly in performance when dimension $d$ grows. Many stochastic programming problems occurring in practice deal with high dimensional random vectors. As argued in [77, Page 3], classical numerical integration techniques, such as the *product-Gauss quadrature*, are only manageable for evaluating integrals of dimension up to 5. Even more sophisticated techniques, such as *lattice methods*, cannot evaluate accurately integrals of dimension greater than 20 (see [77, Page 1]). Another situation that often occurs in practice is when the support of the random vector $\xi$ is finite, but has very large cardinality. In this case, $f(x)$ becomes a weighted sum. Even in that situation it is usually not possible to evaluate $f(x)$ exactly. Indeed, suppose that $\{\xi_i : i = 1, \ldots, d\}$ are independent binary random variables. It follows that the total number of elements of $\operatorname{supp}\xi$ is $2^d$. So, the order of magnitude of the cardinality of the support is $10^{30}$, if $d = 100$. In that case we need to make $10^{30}$ functions evaluations only to calculate $f(x)$ at a single point! This is an astronomically large number from the computational point of view. In such cases it is appealing to resort to sampling techniques.

Here, we consider a standard Monte Carlo approach for replacing problem (1.0.1)

[1] This is usually the case for linear two-stage models with recourse, although the situation changes dramatically for (linear) multistage models.

[2] In principle, the expected value of $F(x,\xi)$ could assume the values $\pm\infty$, for some $x \in X$.

[3] If a closed form formula of function $f$ is known, then there is typically no point in treating problem (1.0.1) as stochastic programming problem. The challenge in solving this kind of problems is that $f$ is not known and it is usually not possible to evaluate it accurately even at a single point $x \in X$.

with a manageable approximation. One considers an independent and identically distributed (i.i.d.) random sample $\{\xi^1, \dots, \xi^N\}$ of $\xi$ and solves the Stochastic Average Approximation (SAA) problem

$$\min_{x \in X} \left\{ \hat{f}_N(x) := \frac{1}{N} \sum_{i=1}^{N} F(x, \xi^i) \right\}. \tag{1.0.2}$$

This approach is commonly known in the stochastic programming literature as the SAA[4] or the *external* sampling approach, but this idea has appeared in the literature under different denominations. For example, it is called the *sample-path* optimization method in [52, 56], whereas in [61] it is called the *stochastic counterpart* method. Let us point out that there exists a variety of sampling-based approaches for solving stochastic programming problems. In [30] the authors present a recent survey on Monte Carlo sampling-based methods for stochastic programming problems and consider, beyond the standard SAA method, other approaches like the Stochastic Approximation (SA) method [55] and variations, the Stochastic Decomposition method of Higle and Sen [28], variance reduction techniques [4, 15, 38, 45], to mention a few. In [40] the authors make an empirical study of the behavior of sampling-based methods for solving stochastic programming problems. Here we focus only on the standard SAA method.

Given a sample realization $\{\xi^1, \dots, \xi^N\}$ of $\xi$, problem (1.0.2) can be seen either as a deterministic optimization problem or as a stochastic programming problem with the empirical probability distribution

$$\hat{\mathbb{P}}(B) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi^i}(B), \forall B \in \mathcal{B}(\mathbb{R}^d), \tag{1.0.3}$$

where $\delta_y(\cdot)$ gives total mass at the point $y \in \mathbb{R}^d$. Indeed, let $\hat{\xi}$ be a random variable having the empirical probability distribution above, it follows that

$$\hat{\mathbb{E}} F(x, \hat{\xi}) = \frac{1}{N} \sum_{i=1}^{N} F(x, \xi^i), \forall x \in X. \tag{1.0.4}$$

So, either suitable algorithms for deterministic optimization problems or also algorithms for stochastic programming problems can be used for solving problem (1.0.2). However, note that the particular algorithm used to solve problem (1.0.2) goes beyond the conceptual framework of the SAA approach and such algorithms are not discussed here. Many interesting questions already arise regarding, for example, in what sense problem (1.0.2) approximates the true problem (1.0.1). In order to pose those questions, let us introduce some basic notation.

[4]To the best of our knowledge, this term was first coined in [37].

Denote the optimal values of problems (1.0.1) and (1.0.2), respectively, by $f^*$ and $\hat{f}_N^*$. Note that $\hat{f}_N^*$ depends on the random sample $\{\xi^1, \ldots, \xi^N\}$, and therefore is a random variable. This quantity can be seen as a statistical estimator of $f^*$. We can also consider an optimal solution $\hat{x}_N$ and the set of optimal solutions $\hat{S}_N$ of problem (1.0.2). These random quantities are also statistical estimators of $x^*$ and $S$, respectively, where $S$ is the set of optimal solutions of problem (1.0.1) and $x^* \in S$. Many questions involving how well these statistical estimators approximate their true counterparts were already answered in the literature. In the sequel we present an extensive overview of these results. In the end of this chapter, we show some new results for an important class of risk averse stochastic programming problems that have not been addressed yet. Before presenting our results, let us begin with the review of the literature.

A natural question that arises is under which conditions the sequences of SAA estimators are (strongly) consistent with respect to their true counterparts. Considering different regularity conditions on the problem data, the strong consistency of sequences of SAA estimators was shown in [21, 36, 56] following the epiconvergence approach. In [73, Section 5.1.1] the consistency of the SAA estimators is derived in an accessible way. Let us summarize the main results presented there. Supposing that w.p.1 $\hat{f}_N(x) \to f(x)$ uniformly in $x \in X$, as $N \to \infty$, it follows immediately that $\hat{f}_N^* \to f^*$ w.p.1 (see [73, Proposition 5.2]). It is worth mentioning that if we only assume that $f(x) = \mathbb{E}F(x, \xi)$ is finite, for every $x \in X$, and that $\{\xi^j : j \in \mathbb{N}\}$ are independent copies of $\xi$, then the strong law of large numbers (see [22, Theorem 2.4.1]) implies the pointwise convergence of $\hat{f}_N$ to $f$, that is: for every $x \in X$ w.p.1[5] $\hat{f}_N(x) \to f(x)$, as $N \to \infty$. However, it is not possible to obtain the uniform convergence only assuming these two conditions.[6]. What is needed is an *uniform* strong law of large numbers[7] for establishing this result. In [73, Theorem 7.55] sufficient conditions on the problem data are presented for guaranteeing that $\hat{f}_N(x)$ converges to $f(x)$ w.p.1 uniformly on $X$. A set of sufficient conditions is: (i) the compactness of the (nonempty) feasible set $X$, (ii) the convexity of the functions $F(\cdot, \xi)$, for $\xi$ in a set of probability 1, (iii) the finiteness of $f(x)$, for every $x$ in a neighborhood of $X$, and (iv) the strong law of large numbers $\mathbb{P}\left[\hat{f}_N(x) \to f(x), \text{ as } N \to \infty\right] = 1$

[5]This means that there exists $A_x \in \mathcal{F}$ such that $\mathbb{P}(A_x) = 1$ and $\hat{f}_N(x, \xi^1(\omega), \cdots, \xi^N(\omega)) \to f(x)$, as $N \to \infty$, for all $\omega \in A_x$. Of course, we suppose that the sequence of random vectors $\{\xi^j : j \in \mathbb{N}\}$ is defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

[6]For example, let us assume that $F(x, \xi) := x\xi$, $\xi \overset{d}{\sim}$ Gaussian$(0, 1)$, and $X = \mathbb{R}$. Note that $f(x) = \mathbb{E}F(x, \xi) = 0$, for all $x \in \mathbb{R}$, and $\hat{f}_N(x) = x\bar{\xi}_N$, where $\bar{\xi}_N := (1/N)\sum_{i=1}^N \xi^i \overset{d}{\sim}$ Gaussian$(0, 1/N)$. We claim that $\hat{f}_N(x)$ does not converge to $f(x) = 0$ uniformly on $x \in \mathbb{R}$. Indeed, take any $\epsilon > 0$. Note that $\mathbb{P}\left[\sup_{x \in \mathbb{R}} \left|\hat{f}_N(x)\right| \geq \epsilon\right] = \mathbb{P}\left[\bar{\xi}_N \neq 0\right] = 1$, for all $N \in \mathbb{N}$. Nevertheless, we have that w.p.1 $\hat{f}_N(x) \to 0$, for every $x \in \mathbb{R}$.

[7]In that case, there exists $A \in \mathcal{F}$ (that does not depend on $x \in X$) that has probability 1 such that $\hat{f}_N(x, \xi^1(\omega), \cdots, \xi^N(\omega)) \to f(x)$ uniformly on $x \in X$, as $N \to \infty$, for every $\omega \in A$.

holds, for every $x$ in a neighborhood of $X$[8]. The key result for proving [73, Theorem 7.55] is the characterization of epiconvergence of functions (see [73, Theorem 7.31]). See also [73, Theorem 7.53] for a different set of assumptions guaranteeing that $\hat{f}_N$ converges uniformly w.p.1 to $f$ on X. In this approach it is not assumed the convexity of $F(\cdot, \xi)$, but its continuity, for $\xi$ in a set of probability 1. This addresses the strong consistency of the SAA optimal values estimators. In order to establish consistency of the SAA estimators of optimal solutions, slightly stronger conditions are assumed (see [73, Theorem 5.3 and 5.4]). In these theorems it was proved that $\mathbb{D}(\hat{S}_N, S) \to 0$ w.p.1. as $N \to \infty$, where

$$\mathbb{D}(A, B) := \sup_{x \in A} \operatorname{dist}(x, B)$$

is the deviation of set $A$ from $B$.

Given the consistency of the SAA estimators, it makes sense to analyze at which rate they converge to their true counterparts values. This analysis was carried out in many publications and can be divided into two types: (a) asymptotic results and (b) non-asymptotic results. The former is related to large sample theory[9], whereas the later obtains results that are also valid in small samples.

The asymptotic results were derived in [63, 64, 65] and in [35] following two different approaches. Assuming some regularity conditions, the asymptotics of the SAA *optimal values* estimators were derived in [64]. The asymptotics of the SAA *optimal solutions* were derived in [65]. In [66] the asymptotics of both the optimal values and the optimal solutions are presented (see also [73, Sections 5.1.2 and 5.1.3]). Here we mention just some results.

Concerning the SAA optimal values estimators, it was shown under some regularity conditions that

$$N^{1/2}(\hat{f}_N^* - f^*) \xrightarrow{d} \inf_{x \in S} Y(x), \tag{1.0.5}$$

where $Y : (\Omega, \mathcal{F}) \to (C(X), \mathcal{B}(C(X)))$ is a random element taking values on the set of continuous functions on (the compact set) $X$ equipped with the sup-norm:

$$\|\phi - \psi\|_{\sup} := \sup_{x \in X} |\phi(x) - \psi(x)|, \ \forall \phi, \psi \in C(X). \tag{1.0.6}$$

Given any $x_1, \ldots, x_k \in X$, the finite-dimensional distribution of the random element $Y$ is given by:

$$(Y(x_1), \ldots, Y(x_k)) \overset{d}{\sim} N\left(0, \Sigma\left(x_1, \ldots, x_k\right)\right), \tag{1.0.7}$$

[8]This condition is superfluous if the sequence of random vectors $\{\xi^i : i \in \mathbb{N}\}$ is taken i.i.d. with $\xi^1 \overset{d}{\sim} \xi$. Indeed, this condition follows immediately from condition (iii) by the strong law of large numbers. However, condition (iv) is used as a way for relaxing the i.i.d. assumption of the random sample. This is particularly useful in order to derive consistency results for nonstandard Monte Carlo sampling-based methods (see, e.g., [29]).

[9]In this setting, the results are obtained by letting the sample size $N$ tend to infinity.

where $\Sigma_{ij} = cov\left[F(x_i,\xi), F(x_j,\xi)\right]$, for all $1 \leq i,j \leq k$[10]. When $S$ is a singleton, it follows from (1.0.5) and (1.0.7) that $\hat{f}_N^*$ has asymptotically normal distribution with mean $f^*$ and variance $\mathrm{Var}\left[F(\bar{x},\xi)\right]/N$, where $\bar{x}$ is the unique solution of problem (1.0.1). This addresses the convergence in distribution of the SAA optimal value estimators. It was also proved that

$$\lim_{N\to\infty} \mathbb{P}\left[N^{1/2}\left(\inf_{x\in S} \hat{f}_N(x) - \hat{f}_N^*\right) \geq \epsilon\right] = 0, \tag{1.0.8}$$

for any $\epsilon > 0$. These results are based on the functional central limit theorem ([2, Corollary 7.17]) and on the delta method in Banach spaces (see [73, Section 7.2.8]). The delta method was used for obtaining a first order approximation of the optimal value function defined on $(C(X), \|\cdot\|_{\sup})$

$$V(g) := \inf_{x\in X} g(x), \forall g \in C(X). \tag{1.0.9}$$

The following conditions are assumed for deriving results (1.0.5) and (1.0.8): (a) $X \subseteq \mathbb{R}^n$ is a nonempty compact set, (b) the sequence of random vectors $\{\xi^i : i \in \mathbb{N}\}$ is i.i.d. with $\xi^1 \overset{d}{\sim} \xi$, (c) $F(x,\xi)$ has finite variance, for every $x \in X$ and (d) there exists a measurable function $\chi : \operatorname{supp}\xi \to \mathbb{R}_+$ having finite second moment such that:

$$|F(x',\xi) - F(x,\xi)| \leq \chi(\xi)\,\|x' - x\| \tag{1.0.10}$$

for every $x', x \in X$ and w.p.1 $\xi$. We will not discuss here the asymptotics of the SAA optimal solutions carried out in [65]. Let us just point out that this analysis was done assuming further regularity conditions on the problem data and by using a second order expansion (second order delta method) of the optimal value function (1.0.9). For more details about this topic one should consult [73, Section 5.1.3] or [66]. Before presenting the non-asymptotic results, let us just mention that the results obtained in [35] are based on a generalized implicit function theorem.

Now we discuss non-asymptotic results for the SAA estimators. This type of results is particularly useful since the involved estimates are valid for any sample size $N \in \mathbb{N}$. Before proceeding, let us introduce some notation. Given $\epsilon \geq 0$, we denote the set of $\epsilon$-optimal solutions of problems (1.0.1) and (1.0.2) by $S^\epsilon$ and $\hat{S}_N^\epsilon$, respectively. When $\epsilon = 0$, we drop the superscripts in $S^\epsilon$ and $\hat{S}_N^\epsilon$ and just write $S$ and $\hat{S}_N$, respectively. Note also that:

$$\begin{aligned} S^\epsilon &= \{x \in X : f(x) \leq f^* + \epsilon\}, \text{ and} & (1.0.11)\\ \hat{S}_N^\epsilon &= \{x \in X : \hat{f}_N(x) \leq \hat{f}_N^* + \epsilon\}. & (1.0.12) \end{aligned}$$

Making use of some results of large deviations theory it is possible to show under some regularity conditions that

$$\mathbb{P}\left(\left[\hat{S}_N^\delta \subseteq S^\epsilon\right] \cap \left[\hat{S}_N^\delta \neq \emptyset\right]\right) \geq 1 - C\exp\left\{-N\beta\right\}, \tag{1.0.13}$$

[10] We drop the dependence of $\Sigma$ on $x_1, \ldots, x_k$ for notational simplicity.

where $C$ and $\beta$ are positive real numbers and $\epsilon > 0$ and $0 \leq \delta < \epsilon$ can be taken arbitrarily. In fact, a bound like (1.0.13) was obtained in [74] for $\epsilon = \delta = 0$ assuming that the problem data satisfies "nice" regularity conditions. In some sense the assumed regularity conditions are not very restrictive, since they are satisfied by two-stage linear stochastic programming problems with a finite number of scenarios[11]. In the sequel we will see how $C$ and $\beta$ depend on the problem data and on the parameters $\epsilon$ and $\delta$ under different regularity conditions. Now, let us point out that estimate (1.0.13) is valid for any sample size $N$ and it shows that the probability of the event "$\hat{S}^{\delta}_{N} \subseteq S^{\epsilon}$" approaches 1 exponentially fast with the increase of the sample size $N$. Note that the event

$$\left[\hat{S}^{\delta}_{N} \subseteq S^{\epsilon}\right] \cap \left[\hat{S}^{\delta}_{N} \neq \emptyset\right] \tag{1.0.14}$$

means that the set of $\delta$-optimal solutions of the SAA problem is nonempty and is contained in the set of $\epsilon$-optimal solution of the "true" problem. If $\epsilon = \delta = 0$, then every optimal solution of the SAA problem is an (exact) optimal solution of the true problem whenever the event (1.0.14) occurs.

Although it is quite remarkable to obtain this type of result for the case $\epsilon = \delta = 0$, in more general situations one is just able to obtain (or is already satisfied in obtaining) an approximate optimal solution of an optimization problem, instead of an exact optimal solution. This analysis can be done by taking $\epsilon > 0$ and $0 \leq \delta \leq \epsilon$. Let us recall that in the SAA approach the optimizer solves (maybe approximately) the SAA problem and hopes that its solution is a good approximation of the optimal solution of the true problem. So, estimate (1.0.13) can be seen as a lower bound of the likelihood of the success of the SAA approach. Equation (1.0.13) can also be used for obtaining an estimate of the sample size $N$ in order to guarantee that the probability of the event (1.0.14) is at least $1-\theta$, where $\theta \in (0,1)$ is a desired value[12]. Indeed, making the right-side of (1.0.13) to be greater than or equal to $1-\theta$ and solving for $N$, we obtain the following estimate for the sample size

$$N \geq \frac{1}{\beta} \log\left(\frac{C}{\theta}\right). \tag{1.0.15}$$

This is an estimate of the *sample complexity* of the SAA approach. In Chapter 2 we define precisely this concept in terms of the parameters $\delta, \epsilon$ and $\theta$.

Now, let us give more details about the results obtained in the literature regarding estimates like (1.0.13). To the best of our knowledge, the first reference that established sufficient conditions for the exponential rate of convergence of the event (1.0.14) is [74]. In that reference the authors analyzed the case $\epsilon = \delta = 0$. Consider

[11]In fact, I dare say that this is the most common type of stochastic programming problem solved in real-life applications.

[12]Usually, $\theta << 1$, like $\theta = 0.10$ or $0.05$ or even $0.01$.

the following assumptions: (a) $f(x)$ is well-defined and finite, for every $x \in X$, (b) $F(\cdot, \xi)$ is convex, for every $\xi \in \operatorname{supp} \xi$, (c) $X$ is closed and convex, (d) $\{\xi^i : i \in \mathbb{N}\}$ are independent random copies of $\xi$, and (e) there exists $c > 0$ and $x^* \in X$ such that

$$f(x) \geq f(x^*) + c \|x - x^*\|, \ \forall x \in X. \tag{1.0.16}$$

Observe that condition (e) implies that $S = \{x^*\}$ and $f^* = f(x^*)$. Assuming conditions (a)-(e), the authors proved that w.p.1 for $N$ large enough

$$\hat{S}_N = S. \tag{1.0.17}$$

Note that since $S$ is a singleton, the events $\left[\hat{S}_N = S\right]$ and $\left[\hat{S}_N \subseteq S\right] \cap \left[\hat{S}_N \neq \emptyset\right]$ are the same. Equivalently, they have shown that the event

$$\liminf_{N\to\infty} \left[\hat{S}_N = S\right] := \bigcup_{M\in\mathbb{N}} \bigcap_{N\geq M} \left[\hat{S}_N = S\right] \tag{1.0.18}$$

has probability 1. Of course, the index $M \in \mathbb{N}$ such that $\left[\hat{S}_N = S\right]$, for every $N \geq M$ depends on the sequence realization $\{\xi^i(\omega) : i \in \mathbb{N}\}$. Assuming additionally that $\operatorname{supp} \xi$ is finite[13], they have shown that there exist $C > 0$ and $\beta > 0$ such that

$$\mathbb{P}\left(\left[\hat{S}_N \subseteq S\right] \cap \left[\hat{S}_N \neq \emptyset\right]\right) \geq 1 - C\exp\{-N\beta\}, \forall N \in \mathbb{N}. \tag{1.0.19}$$

In fact, it was proved that there exists $\beta > 0$ such that

$$\limsup_{N\to\infty} \frac{1}{N} \log\left(\mathbb{P}\left[\hat{S}_N \neq S\right]\right) < -\beta, \tag{1.0.20}$$

which is equivalent to (1.0.19), see, for example, Proposition 2.8.7. Note also that since

$$\mathbb{P}\left[\hat{S}_N \neq S\right] \leq C\exp\{-N\beta\} \tag{1.0.21}$$

we have that

$$\sum_{N\in\mathbb{N}} \mathbb{P}\left[\hat{S}_N \neq S\right] \leq \frac{C\exp\{-N\beta\}}{1 - \exp\{-N\beta\}} < \infty. \tag{1.0.22}$$

It follows from the Borel-Cantelli Lemma (see [22, Theorem 2.3.1]) that

$$\limsup_{N\to\infty} \left[\hat{S}_N \neq S\right]$$

is a set of null probability, equivalently

$$\liminf_{N\to\infty} \left[\hat{S}_N = S\right] = \left(\limsup_{N\to\infty} \left[\hat{S}_N \neq S\right]\right)^C$$

[13] Indeed, for showing this result the authors assumed an even less restrictive condition than the finite number of scenarios (see [74, Assumption B] or [73, Page 192, assumption (M4)]).

is a set of probability[14] 1.

In the same reference, the authors have shown a similar result without assuming that $S$ is a singleton. For proving this result, they assumed conditions (a) and (d) as before, and: (b') $F(\cdot, \xi)$ is a polyhedral function (see Definition 2.5.19), for every $\xi \in \operatorname{supp} \xi$, (c') $X$ is a polyhedron (see Definition 2.5.4), (e') there exists $c > 0$ and $A \subseteq X$ nonemepty bounded such that

$$f(x) \geq f(x^*) + c \operatorname{dist}(x, A), \tag{1.0.23}$$

for any $x \in X$ and $x^* \in A$. Condition (e') implies that $S = A$. Assuming additionally that $\operatorname{supp} \xi$ is finite, the authors proved that $S$ is a compact polyhedron and that there exists positive real numbers $C$ and $\beta$ such that

$$\mathbb{P}\left(\left[\hat{S}_N \text{ is a face of set } S\right] \cap \left[\hat{S}_N \neq \emptyset\right]\right) \geq 1 - C\exp\{-N\beta\}, \ \forall N \in \mathbb{N}. \tag{1.0.24}$$

Note that the event $\left[\hat{S}_N \text{ is a face of set } S\right]$[15] is contained in the event $\left[\hat{S}_N \subseteq S\right]$, which implies that the lower bound (1.0.24) is also valid for the event (1.0.14). Moreover, again by the Borel-Cantelli Lemma, we also conclude that w.p.1 for $N$ large enough the event (1.0.14) occurs. It is worth mentioning that two-stage linear stochastic programming problems with a finite number of scenarios satisfy all these conditions (see also [73, Section 5.3.3]) provided that $f$ is proper and $S$ is nonempty and bounded.

In [37] the authors analyzed how well SAA estimators approximate their true counterparts for discrete stochastic programming problems, i.e. assuming that the feasible set $X$ is finite. Assuming additionally that: (a) for any $x \in X$, the random variable $F(x, \xi)$ has finite expected value $f(x)$, and (b) $\{\xi^i : i \in \mathbb{N}\}$ are independent copies of $\xi$, they proved that for any given numbers $\epsilon \geq 0$ and $0 \leq \delta \leq \epsilon$, the following statements hold true:

(i) w.p.1 $\hat{f}^*_N \to f^*$, as $N \to \infty$.

(ii) w.p.1 for $N$ large enough $\hat{S}^\delta_N \subseteq S^\epsilon$.

The key step for proving items (i) and (ii) is to show that w.p.1. $\hat{f}_N(x) \to f(x)$ uniformly in $x \in X$. This result follows from the finiteness of $X$ and the fact that w.p.1 $\hat{f}_N(x) \to f(x)$, as $N \to \infty$, by the strong law of large numbers. Now, items (i) and (ii) follow easily[16].

[14] This remark shows the well-known fact that results like (1.0.19) are stronger than just showing that w.p.1 for $N$ large enough $\left[\hat{S}_N = S\right]$.

[15] In Definition 2.5.3 we recall the definition of a face of a convex set.

[16] Let us point out that the finiteness of $X$ also guarantees that $S \neq \emptyset$ and that w.p.1 $\hat{S}_N \neq \emptyset$, for all $N \in \mathbb{N}$. So, we also have that w.p.1 for $N$ large enough $\left[\hat{S}^\delta_N \subseteq S^\epsilon\right] \cap \left[\hat{S}^\delta_N \neq \emptyset\right]$ happens.

Rates of convergence for the SAA estimators were also derived. Here we present the result for the SAA optimal solutions. Take any $\epsilon \geq 0$ and $0 \leq \delta \leq \epsilon$. If $X \backslash S^\epsilon = \emptyset$, then $\hat{S}_N^\delta \subseteq X = S^\epsilon$ immediately. So, take $X \backslash S^\epsilon \neq \emptyset$. Assuming that: (c) there exists $\mathrm{u} : X \backslash S^\epsilon \to X$ such for each $x \in X \backslash S^\epsilon$ the random variable $Z_x := F(\mathrm{u}(x), \xi) - F(x, \xi)$ has finite moment generating function (henceforth called m.g.f.) in a neighborhood of 0; they proved that

$$\mathbb{P}\left[\hat{S}_N^\delta \subseteq S^\epsilon\right] \geq 1 - \mathrm{card}\left(X \backslash S^\epsilon\right) \exp\left\{-N\beta\right\}, \; \forall N \in \mathbb{N}, \tag{1.0.25}$$

where u satisfies $f(\mathrm{u}(x)) < f(x) - \epsilon$, for all $x \in X \backslash S^\epsilon$, and $\beta = \beta(\epsilon, \delta) > 0$ is given by:

$$\beta := \min_{x \in X \backslash S^\epsilon} I_x(-\delta). \tag{1.0.26}$$

The function $I_x(\cdot)$ is known in the large deviation theory as a rate function. This particular rate function is defined as the convex conjugate of the log-moment generating function of the random variable $Z_x$, that is:

$$I_x(s) := \left(\log M_x\right)^* (s) = \sup_{t \in \mathbb{R}} \{st - \log M_x(t)\},$$

where

$$M_x(t) := \mathbb{E} \exp\left\{tZ_x\right\}.$$

Let us give more details about some points related to the derivation of the exponential rate of convergence (1.0.25). First, note that there exists a function $\mathrm{u} : X \backslash S^\epsilon \to X$ satisfying $f(\mathrm{u}(x)) - f(x) < \epsilon$, for all $x \in X \backslash S^\epsilon$. Indeed, we can always define $\mathrm{u}(x) \in S$, for any $x \in X \backslash S^\epsilon$. Then,

$$f(x) > f^* + \epsilon = f(\mathrm{u}(x)) + \epsilon, \forall x \in X \backslash S^\epsilon. \tag{1.0.27}$$

Second, note that the assumption (c) is weaker than the following: $\forall x' \in S$, $\forall x \in X \backslash S$, the m.g.f. $M_{x',x}(t)$ of the random variable $Z_{x',x} := F(x', \xi) - F(x, \xi)$ is finite for $t$ in a neighborhood of 0. Third, observe that

$$\mathbb{E} Z_x = f(\mathrm{u}(x)) - f(x) \leq - \min_{x \in X \backslash S^\epsilon} \left(f(x) - f(\mathrm{u}(x))\right) =: -\epsilon^* < -\epsilon, \tag{1.0.28}$$

since $X \backslash S^\epsilon$ is finite. Now, let us obtain a lower bound for $\beta = \beta(\epsilon, \delta) > 0$ that is valid for any $\epsilon \geq 0$ sufficiently small and any $0 \leq \delta \leq \epsilon$. We begin by noting that since $M_x(\cdot)$ is finite in a neighborhood of zero, we have that

$$I_x(s) = \frac{\left(s - \mathbb{E} Z_x\right)^2}{2\sigma_x^2} + \mathrm{o}\left(|s - \mathbb{E} Z_x|^2\right) \tag{1.0.29}$$

for $s$ in a neighborhood of $\mathbb{E} Z_x = f(\mathrm{u}(x)) - f(x)$, where $\sigma_x^2 := \mathrm{Var}[Z_x]$ for all $x \in X \backslash S^\epsilon$. Using again that $X \backslash S^\epsilon$ is finite, we obtain that there exists $\epsilon_0 > 0$ such that

$$I_x(s) \geq \frac{\left(s - \mathbb{E} Z_x\right)^2}{3\sigma_x^2}, \tag{1.0.30}$$

for $\mathbb{E}Z_x \leq s \leq \mathbb{E}Z_x + \epsilon_0$ and for all $x \in X\backslash S^\epsilon$. Taking $\epsilon_0 > 0$ smaller if necessary, we can assume without loss of generality that $\epsilon_0 \leq \min_{x\in X\backslash S}(f(x) - f^*)$. We claim that for any $0 \leq \epsilon < \epsilon_0$ and any $0 \leq \delta \leq \epsilon$, the following inequality is satisfied

$$I_x(-\delta) \geq \frac{(\epsilon_0 - \delta)^2}{3\sigma^2_{\max}}, \forall x \in X\backslash S^\epsilon, \tag{1.0.31}$$

where $\sigma^2_{\max} := \max_{x\in X\backslash S^\epsilon} \sigma^2_x$. Before showing that inequality (1.0.31) is satisfied, observe that it gives the following lower bound for $\beta \geq (\epsilon_0 - \delta)^2/3\sigma^2_{\max} > 0$. Using this lower bound in (1.0.25) we obtain:

$$\mathbb{P}\left[\hat{S}^\delta_N \subseteq S^\epsilon\right] \geq 1 - \mathrm{card}\,(X\backslash S^\epsilon) \exp\left\{-\frac{N(\epsilon_0 - \delta)^2}{3\sigma^2_{\max}}\right\} \tag{1.0.32}$$

for any $\epsilon \geq 0$ sufficiently small and $0 \leq \delta \leq \epsilon$. For showing that (1.0.31) holds, let us first observe that $\mathbb{E}Z_x \leq -\epsilon_0 < -\epsilon \leq -\delta$, for any $x \in X\backslash S^\epsilon$. If $-\delta \in [\mathbb{E}Z_x, \mathbb{E}Z_x + \epsilon_0]$, then

$$I_x(-\delta) \geq \frac{(-\delta - \mathbb{E}Z_x)^2}{3\sigma^2_x} \geq \frac{(\epsilon_0 - \delta)^2}{3\sigma^2_{\max}}, \tag{1.0.33}$$

since $\sigma^2_{\max} \geq \sigma^2_x$ and $0 \leq \epsilon_0 - \delta \leq -\delta - \mathbb{E}Z_x$. If $-\delta > \mathbb{E}Z_x + \epsilon_0$, then using the fact that $I_x(\cdot)$ is monotone in $[\mathbb{E}Z_x, \infty)$ we obtain that

$$I_x(-\delta) \geq I_x(\mathbb{E}Z_x + \epsilon_0) \geq \frac{\epsilon^2_0}{3\sigma^2_x} \geq \frac{(\epsilon_0 - \delta)^2}{3\sigma^2_{\max}}. \tag{1.0.34}$$

For finishing the review of [37] regarding properties of SAA estimators for discrete problems, let us just say that asymptotic results for the SAA optimal value estimator were derived using the central limit theorem.

In [75] the authors derived sample complexity estimates for general static or two-stage stochastic programming problems. There, they do not suppose that $X$ is finite or that the random data $\xi$ has finite support or even that w.p.1 $F(\cdot, \xi)$ is convex. Instead, they suppose that $X$ has finite diameter and that the random costs $\{F(x, \xi) : x \in X\}$ do not present "wild" randomness. The results were derived using the large deviation theory, in particular the upper bound of Cramer's large deviation theorem. For such, one must suppose the finiteness (in a neighborhood of 0) of the m.g.f. of the involved random variables. Moreover, they suppose that the differences $F(x, \xi)-F(x', \xi)$, for all $x, x' \in X$, also do not present "wild" randomness and that the closer $x$ and $x'$ are to each other, the higher is the probability that $F(x, \xi) - F(x', \xi)$ is close to 0. For the record, let us enumerate precisely these assumptions: (a) $f(x) = \mathbb{E}F(x, \xi)$ is finite, for any $x \in X$, (b) the (nonempty) feasible set $X$ has finite diameter $D$, (c) $\{\xi^i : i \in \mathbb{N}\}$ are independent copies of $\xi$,

(d) for all $x, x' \in X$, the random variable $(F(x,\xi) - F(x',\xi)) - (f(x) - f(x'))$ has m.g.f. $M_{x',x}(t)$ satisfying[17]

$$M_{x',x}(t) \leq \exp\left\{\frac{1}{2}\sigma^2 t^2\right\}, \ \forall t \in \mathbb{R}, \tag{1.0.35}$$

where $\sigma > 0$ is a finite constant, and (e) there exists a measurable function $\chi : \operatorname{supp}\xi \to \mathbb{R}_+$ that has finite m.g.f. in a neighborhood of 0, and that satisfies[18]

$$|F(x,\xi) - F(x',\xi)| \leq \chi(\xi) \|x - x'\|, \tag{1.0.36}$$

for all $x, x' \in X$ and all $\xi \in \operatorname{supp}\xi$. Assuming conditions (a)-(e), in [75, Theorem 1] the authors obtained an estimate of the sample size $N$ that guarantees that with probability at least $1-\theta$ every $\delta$-optimal solution of the SAA problem is an $\epsilon$-optimal solution of the true problem (see [75, Equation (22)]). A similar derivation is obtained in [73, Theorem 5.18] that we present below. Take any $\epsilon > 0$ and $0 \leq \delta < \epsilon$. It was shown that

$$\mathbb{P}\left[\hat{S}^\delta_N \subseteq S^\epsilon\right] \geq 1 - \exp\{-N\beta\} - \left(\frac{4\rho\tilde{M}D}{\epsilon - \delta}\right)^n \exp\left\{-\frac{N(\epsilon-\delta)^2}{8\sigma^2}\right\} \tag{1.0.37}$$

for all $N \in \mathbb{N}$. Let us make some remarks about the quantities appearing in equation (1.0.37). First, $\tilde{M}$ can be taken as any number greater than $M := \mathbb{E}\chi(\xi) < +\infty$ that is finite since $M_\chi(t)$ is finite in a neighborhood of 0 by hypothesis (e) (see also Proposition 2.8.6). The constant $\beta$ is equal to $I_\chi(\tilde{M}) \in (0, +\infty]$ that we know is positive (maybe even equal to $+\infty$), since $\tilde{M} > \mathbb{E}\chi(\xi)$ (see also Remark 2.1.7). Finally, $\rho$ is a universal constant that is related to $v$-nets in $\mathbb{R}^n$ (see Definition 2.8.2). Let us assume that the diameter of $X$ is a positive real number[19]. It is a well-known fact that there exists a positive absolute constant $\rho$[20] such that for every $0 < v \leq D := \operatorname{diam} X$ there exists a $v$-net $\tilde{X}$ of $X$ such that

$$\operatorname{card}\tilde{X} \leq \left(\frac{\rho D}{v}\right)^n. \tag{1.0.38}$$

The proof of [73, Theorem 5.18] (or equivalently [75, Theorem 1]) used the sample complexity estimate obtained in [37] for discrete stochastic programming problems for deriving an estimate for general stochastic problems with bounded feasible sets.

[17] We say that a random variable satisfying (1.0.35) is a $\sigma$-sub-Gaussian random variable (see Section 2.4).

[18] More precisely, they have supposed that inequality (1.0.36) is satisfied with right-side equal to $\chi(\xi)\|x - x'\|^\gamma$, where $\gamma > 0$. Here, we just present the case $\gamma = 1$.

[19] If $D = 0$, then $X$ is a singleton and we always have that $\hat{S}^\delta_N \subseteq S^\epsilon = X$.

[20] In Proposition 2.8.3 we proved that $\rho$ can be taken less than or equal to 5.

Given any $\epsilon > 0$ and $0 \leq \delta < \epsilon$, take any $\tilde{M} > M$ and consider a $v$-net $\tilde{X}$ of $X$, satisfying

$$\operatorname{card} \tilde{X} \leq \left( \frac{4\rho D \tilde{M}}{\epsilon - \delta} \right)^n,$$

where

$$v := \frac{\epsilon - \delta}{4\tilde{M}} > 0.$$

Now, one can consider the true and the SAA problems restricted to the net $\tilde{X}$, instead of the original set $X$. Considering appropriate parameters $\delta < \delta' < \epsilon' < \epsilon$, one can derive the following estimate

$$\begin{aligned} \mathbb{P}\left[ \tilde{S}_N^{\delta'} \subset \tilde{S}^{\epsilon'} \right] &\geq 1 - \operatorname{card}\left( \tilde{X} \backslash \tilde{S}^{\epsilon'} \right) \exp\left\{ -\frac{N(\epsilon' - \delta')^2}{2\sigma^2} \right\} &(1.0.39)\\ &\geq 1 - \left( \frac{4\rho \tilde{M} D}{\epsilon - \delta} \right)^n \exp\left\{ -\frac{N(\epsilon' - \delta')^2}{2\sigma^2} \right\} &(1.0.40) \end{aligned}$$

using the estimate derived for discrete stochastic programming problems (this gives the estimate of the last term of the right-side of (1.0.37) by considering $\epsilon' - \delta' = (\epsilon - \delta)/2)$). Furthermore, since $I_\chi(t)$ is finite for $t$ in a neighborhood of 0 and $\tilde{M} > \mathbb{E}\chi(\xi)$, the upper bound of the Cramer's large deviation theorem implies that

$$\mathbb{P}\left[ \frac{1}{N} \sum_{i=1}^N \chi(\xi^i) > \tilde{M} \right] \leq \exp\{-N\beta\}, \tag{1.0.41}$$

where $\beta = I_\chi(\tilde{M}) > 0$. It is also readily seen that $\hat{f}_N$ is $\left( \frac{1}{N} \sum_{i=1}^N \chi(\xi^i) \right)$-Lipschitz continuous in $X$. So, whenever the event $\left[ \frac{1}{N} \sum_{i=1}^N \chi(\xi^i) \leq \tilde{M} \right]$ happens, we have that $\hat{f}_N$ is $\tilde{M}$-Lipschitz continuous in $X$. Noting also that $f$ is $M$-Lipschitz in $X$, one can show that the event $\left[ \hat{S}_N^\delta \subset S^\epsilon \right]$ happens, whenever the event

$$\left[ \tilde{S}_N^{\delta'} \subset \tilde{S}^{\epsilon'} \right] \bigcap \left[ \frac{1}{N} \sum_{i=1}^N \chi(\xi^i) \leq \tilde{M} \right] \tag{1.0.42}$$

occurs. Observe that the right-side of equation (1.0.37) is just a lower estimate for the probability of the event (1.0.42). It is also possible to prove that, under the assumed regularity conditions, the event $\left[ \hat{S}_N \neq \emptyset \right]$ has probability 1. So, the probability of the event (1.0.14) is equal to the left-side of (1.0.37). That is the general idea of the proof. One should consult [73, Theorem 5.18] for more details. We also derive a similar lower bound in Section 2.1.1 applying directly the uniform exponential bound theorem (see Theorem 2.1.5).

From (1.0.37) it is possible to obtain an estimate of the sample size $N$ that guarantees that with probability at least $1-\theta$, for $\theta \in (0,1)$ given, every $\delta$-optimal solution of the SAA problem is an $\epsilon$-optimal solution of the true problem. One way for achieving this is to derive the minimum value of $N \in \mathbb{N}$ such that both inequalities

$$\begin{aligned} \exp\left\{-NI_\chi(\tilde{M})\right\} &\leq \frac{\theta}{2}, \text{ and} && (1.0.43)\\ \left(\frac{4\rho\tilde{M}D}{\epsilon-\delta}\right)^n \exp\left\{-\frac{N(\epsilon-\delta)^2}{8\sigma^2}\right\} &\leq \frac{\theta}{2} && (1.0.44)\end{aligned}$$

hold true. This gives us the following estimate for $N$

$$N \geq \frac{8\sigma^2}{(\epsilon-\delta)^2}\left[n\log\left(\frac{4\rho\tilde{M}D}{\epsilon-\delta}\right)+\log\left(\frac{2}{\theta}\right)\right]\bigvee\left[\frac{1}{I_\chi(\tilde{M})}\log\left(\frac{2}{\theta}\right)\right], \tag{1.0.45}$$

where $a \vee b := \max\{a,b\}$, for all $a, b \in \mathbb{R}$. Treating $0 \leq \delta < \epsilon$ as varying parameters, we note that the maximum in (1.0.45) is achieved by its first term for sufficiently small values of $\epsilon-\delta > 0$. Moreover, if assumption (e) is satisfied with $\chi(\xi) = M$ for $\xi$ in a set of probability 1, then $I_\chi(\tilde{M}) = +\infty$, for any $\tilde{M} > M$. In that case, the second term in the maximum (1.0.45) vanishes, and it suffices to take

$$N \geq \frac{8\sigma^2}{(\epsilon-\delta)^2}\left[n\log\left(\frac{4\rho MD}{\epsilon-\delta}\right)+\log\left(\frac{1}{\theta}\right)\right]. \tag{1.0.46}$$

Although the authors have pointed out in [75] that sample complexity estimates like (1.0.45) or (1.0.46) are too conservative to be useful in practice, these type of results provide a theoretical guarantee that static or two-stage stochastic programming problems, that are in some sense well-behaved (see assumptions (a)-(e)), can be efficiently approximated by the Monte Carlo sampling-based approach provided that we do not ask for too much accuracy, that is, values of $\epsilon$ that are too small. Indeed, if $\epsilon$ is too small, then estimate (1.0.45) indicates that the SAA approach could be impractical for obtaining an $\epsilon$-optimal solution of the true problem. Another point of concern is if the SAA problem can be efficiently solved. This typically depends also if it is possible to evaluate efficiently the function values $F(x,\xi)$, for any $x \in \mathbb{R}^n$ and $\xi \in \operatorname{supp}\xi$, and/or if the functions $F(\cdot,\xi)$ and the feasible set $X$ are convex[21].

For closing the review of [75], let us just mention that the authors also commented about sample complexity estimates for the multistage setting. They reasoned that although it is possible to write the multistage problem as a static or two-stage problem using the cost-to-go functions, there is an important difference between both

[21] In that case, by calling an oracle one can usually also obtain a subgradient $s(x,\xi) \in \partial_x F(x,\xi)$, for every $x \in \mathbb{R}^n$ and $\xi \in \operatorname{supp}\xi$. Proceeding in that way, one can obtain a polyhedral function $\hat{g}_N(\cdot)$ that approximates $\hat{f}_N(\cdot)$ from below in $X$.

settings. Indeed, for multistage problems it is not possible to evaluate $F(x,\xi)$ accurately. For example, for 3-stage problems, the exact evaluation of $F(x,\xi)$ involves solving *exactly* a two-stage problem with initial conditions determined by $x$ and $\xi$. We are not saying that one needs to evaluate $F(x,\xi)$ exactly in order to obtain a reasonable approximation of the true problem. What we are trying to convey is that, differently from two-stage problems, $F(x,\xi)$ cannot be accurately evaluated for $T$-stage problems ($T \geq 3$) and that this has important consequences in the sample complexity estimates for multistage problems.

In the multistage setting, the SAA approach consists in building a scenario tree (see Section 2.2) using a conditional sampling scheme (see Section 2.1.2). Supposing that for each node of the scenario tree at level $t = 1, \ldots, T-1$, one generates $N_{t+1}$ children nodes, the constructed SAA scenario tree has

$$N = \prod_{t=2}^{T} N_t \tag{1.0.47}$$

number of scenarios. In [67] it was shown that in order to obtain consistency results for the SAA estimators in the multistage setting, one has to make $N_t \to \infty$, for each $t = 2, \ldots, T$. This is a first indicator that in order for the SAA problem approximates arbitrarily well the true problem, the total number of scenarios in the multistage setting must grow exponentially fast with respect to the number of stages $T$.

The analysis of the exponential rates of convergence of the SAA estimators in the multistage setting was carried out in [69]. In this paper the author obtained sample complexity estimates for a 3-stage stochastic programming problem under similar hypothesis considered in the analysis of two-stage problems (see [69] or Section 2.1.2 for more details). In a nutshell, he proves that if one takes $N_2 = N_3$[22] satisfying

$$N_2 \geq \frac{\mathrm{O}(1)\sigma^2}{\epsilon^2}\left[(n_1+n_2)\log\left(\frac{DM}{\epsilon}\right) + \log\left(\frac{\mathrm{O}(1)}{\theta}\right)\right], \tag{1.0.48}$$

then

$$\mathbb{P}\left[\hat{S}^{\epsilon/2}_{N_2,N_3} \subseteq S^{\epsilon}\right] \geq 1-\theta \tag{1.0.49}$$

where $\epsilon > 0$ and $\theta \in (0,1)$ are given. This estimate suggests that the total number of scenarios $\prod_{t=2}^{T} N_t$ in the scenario tree grows at least to order of

$$\left(\frac{\sigma^2(n_1+\cdots+n_{T-1})}{\epsilon^2}\right)^{T-1}. \tag{1.0.50}$$

In Section 2.1.2 we extend this analysis for $T$-stage stochastic programming problems with arbitrary $T \in \mathbb{N}$ obtaining an estimate like (1.0.50).

[22] This point is not important and it is made here only for simplifying the exposition. In [69] this assumption was not used to derive the sample complexity estimates.

Note that all references discussed so far derived results for risk neutral stochastic programming problems. A relevant question that can be posed is: why do we optimize the expected value of the random cost? If one has to repeat the optimization procedure many times under the same initial conditions, then, by the law of large numbers, optimizing the expected value gives an optimal decision in the long run. However, this reasoning does not appeal to all decision making situations. A typical example occurs in the area of portfolio selection where a trade-off between risk and return is taken into account in the optimization problem. For instance, a strategy that invests all wealth in just one asset, the one that has the maximum expected return, appeals to almost nobody, although this is the optimal strategy for the portfolio selection problem considering as the optimization criterion just the expected value operator. The drawback of this strategy is clear, one does not give any importance to deviations of the portfolio return with respect to its expected return [23]. The trade-off between return and risk is the cornerstone of the modern theory of portfolio selection introduced in the pioneering work of [41] and further analyzed in [42, 43]. In the mean-variance models introduced by Markowitz, besides considering the expected portfolio return (or cost) one uses the standard deviation of the portfolio return as a way to penalize deviations from its expected cost. Many theoretical advances were made in this area in recent decades [1, 3, 17, 25, 47, 58] and much attention has been given to risk averse stochastic programming problems [19, 26, 31, 39, 51, 57, 70, 76], to mention a few. In this type of problems, one uses a risk averse risk measure $\mu$ in order to summarize the random cost that depends on the decision $x \in X$ into a real number. A general risk averse static stochastic programming problem is formulated as

$$\min_{x \in X} \left\{ v(x) := \mu\left(F(x, \xi)\right) \right\}, \tag{1.0.51}$$

where $\xi$, $x$, $X$ and $F$ are defined as before (see paragraph after (1.0.1)).

When $\mu$ is a (regular)[24] law invariant[25] risk measure (see Section 2.7), also known as *version independent* risk measure, and it is possible to sample from the random

[23]For example, consider that there are two possible scenarios $\{1, 2\}$ each with equal probability of occurring, and assets $A$ and $B$. An investor wants to allocate all of his wealth on this two assets. Asset $A$ returns $-100\%$ when scenario 1 occurs and $+200\%$ when scenario 2 occurs. Asset $B$ returns 50% in both scenarios. Observe that the expected return of these two assets are equal to 50%. Therefore, the portfolios that invest all wealth in each asset are equally good using the expected return criterion. However, note that investing in asset $A$ is much more risky than investing in asset $B$. Indeed, the investor has 50% of chance of losing all his wealth if he invests only in asset A.

[24]See [73, Definition 6.45].

[25]In a nutshell, a (regular) law invariant risk measure $\mu : L_p(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$, $p \in [1, \infty)$, can be seen as defined on the set of all cumulative distributions functions of $\mathbb{R}$. In that case we can write $\mu(Z) = \mu(F_Z)$, where $F_Z(z) := \mathbb{P}[Z \leq z]$, for all $z \in \mathbb{R}$, is the cumulative distribution function of the random variable $Z$. Given a sample realization $\{Z_1, \ldots, Z_N\}$ of the random variable $Z$, we consider the empirical cumulative distribution function $\hat{F}_N(z) := \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{Z_i \leq z\}}$ that is,

vector $\xi$, one can easily consider Monte Carlo sampling-based approaches to approximate problem (1.0.51) by its empirical counterpart. Given a random sample $\{\xi^i : i = 1, \ldots, N\}$ of $\xi$, consider the following problem

$$\min_{x \in X} \left\{ \hat{v}_N(x) := \hat{\mu}\left(F(x, \hat{\xi})\right) \right\}, \tag{1.0.52}$$

where $F(x, \hat{\xi})$ has the empirical cumulative distribution function

$$\hat{H}_N(x, z) := \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{F(x,\xi^i) \leq z\}}, \text{ for every } z \in \mathbb{R},$$

and $\hat{\mu}\left(F(x, \hat{\xi})\right) := \mu\left(\hat{H}_N(x, \cdot)\right)$. Here we also denote problem (1.0.52) as the SAA problem. Akin to the risk neutral case, we can see the SAA optimal value and the SAA optimal solutions as statistical estimators of their true counterparts. The same questions regarding the: (a) (strong) consistency, (b) asymptotic distribution, and (c) rates of convergence of these statistical estimators can be considered in the risk averse setting.

Let us make a brief review of the literature addressing these questions. [71] is the only reference that we are aware of that studies the convergence of empirical estimates of risk averse stochastic programming problems. Although many other references like [11, 32, 33, 50, 80] have established either the strong consistency and/or the asymptotic distribution of the empirical estimates for some risk measures. However, it is important to differentiate these type of results. In the latter type, one considers as given a risk measure $\mu$ satisfying some regularity conditions and shows that, for a fixed random variable $Z$, its statistical estimator $\hat{\mu}_N(Z)$ converges in some sense to its true counterpart $\mu(Z)$, when $N$ goes to $\infty$. In [71] this kind of result was also proved for a broad class of risk measures, but also that the statistical estimators associated with the risk averse stochastic programming problem, that is, the optimal value estimator $\hat{v}_N^*$ and the set of optimal solutions estimator $\hat{S}_N$ converged in some sense to their true counterparts $v^*$ and $S$, respectively. As it was shown in the risk neutral case, these results follows by proving that $\hat{\mu}_N(F(x, \hat{\xi}))$ converges to $\mu(F(x, \xi))$ uniformly on $x \in X$.

Now we make a brief presentation of the main results in [71]. The author begun by investigating convergence properties of the empirical estimates of law invariant convex risk measures. In Section 2.7 we recall the definition of this type of risk

in fact, a valid cumulative distribution function of $\mathbb{R}$ (see Section 2.3). Therefore, the random quantity $\mu\left(\hat{F}_N\right)$ is the empirical estimate of $\mu(F)$. The expected value operator is an example of law invariant risk measure. We will see many other examples in the sequel. Given a sample realization $\{Z_i : 1 \leq i \leq N\}$ of a random variable $Z$, note that the sample mean $\bar{Z} = \frac{1}{N}\sum_i^N Z_i$ is just the expected value of the random variable $\hat{Z}$ that has cumulative distribution function $\hat{F}_N(z) := \frac{1}{N}\sum_{i=1}^N \mathbb{1}_{\{Z_i \leq z\}}$, i.e., $\bar{Z} = \hat{E}\hat{Z}$.

measures. It is worth mentioning that this is a broad[26] class of risk measures. Take any $Z \in L_p(\Omega, \mathcal{F}, \mathbb{P})$, $p \in [1, \infty)$, and consider a sequence $\{Z^i : i \in \mathbb{N}\}$ of i.i.d. copies of $Z$ defined also on $L_p(\Omega, \mathcal{F}, \mathbb{P})$. Let $F_Z(z) := \mathbb{P}[Z \leq z]$ be the cumulative distribution function of $Z$. Moreover, for every $N \in \mathbb{N}$ and $\omega \in \Omega$, let

$$\hat{F}_N(z, \omega) := \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{Z^i(\omega) \leq z\}} \tag{1.0.53}$$

be the empirical cumulative distribution function associated with the sequence of random variables $\{Z^i : i \in \mathbb{N}\}$. In [71, Theorem 2.1] it was proved that if $\mu : L_p(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$, $p \in [1, \infty)$, is a (finite-valued) law invariant convex risk measure, then w.p.1

$$\mu\left(\hat{F}_N\right) \to \mu\left(F_Z\right), \tag{1.0.54}$$

as $N$ goes to $\infty$. Now consider the risk averse stochastic programming problem

$$\min_{x \in X} \left\{ v(x) := \mu\left(F(x, \xi)\right) \right\}, \tag{1.0.55}$$

where $\xi : \Omega \to \mathbb{R}^d$ is the random data, $F : \mathbb{R}^n \times \mathbb{R}^d \to \bar{\mathbb{R}}$ is the random cost function, $\mu$ is a law invariant risk measure and $X \subseteq \mathbb{R}^n$ is the feasible set. We also consider a sequence $\{\xi^i : i \in \mathbb{N}\}$ of i.i.d. copies of $\xi$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$. For every $x \in \mathbb{R}^n$, $\omega \in \Omega$ and every $N \in \mathbb{N}$ consider

$$\hat{H}_N(x, \omega, \cdot) := \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{F(x, \xi^i(\omega)) \leq \cdot\}} \tag{1.0.56}$$

the empirical distribution function associated with the sequence of random variables $\{F(x, \xi^i) : i \in \mathbb{N}\}$. Define

$$\hat{v}_N(x, \omega) := \mu\left(\hat{H}_N(x, \omega)\right), \ \forall (x, \omega) \in \mathbb{R}^n \times \Omega. \tag{1.0.57}$$

Under appropriate regularity conditions it follows that $v(\cdot)$ is a l.s.c. function and $\hat{v}_N$ epiconverges to $v$ w.p.1. In fact, the following conditions are sufficient for deriving these result (see [71, Theorem 3.1]): (a) $\mu : L_p(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ is a law invariant convex risk measure, (b) $(x, \omega) \in \mathbb{R}^n \times \Omega \to F(x, \xi(\omega))$ is a normal integrand (see Definition 2.6.12), (c) for every $x \in \mathbb{R}^n$, the function $\omega \in \Omega \to F(x, \xi(\omega))$ belongs to $L_p(\Omega, \mathcal{F}, \mathbb{P})$, and (d) for every $\tilde{x} \in \mathbb{R}^n$ there exists a neighborhood $V_{\tilde{x}}$ of $\tilde{x}$ and a function $h \in L_p(\Omega, \mathcal{F}, \mathbb{P})$ such that $F(x, \xi(\cdot)) \geq h(\cdot)$, for every $x \in V_{\tilde{x}}$.

Assuming additionally that $F(\cdot, \xi)$ is a convex function w.p.1 $\xi$, as a consequence of the epiconvergence of $\hat{v}_N(\cdot)$ to $v(\cdot)$, it follows that $\hat{v}_N(x, \omega) \to v(x)$ uniformly on

[26] Every coherent risk measure is also a convex risk measure. It is also true that every Optimized Certainty Equivalent (OCE) risk measure is also a law invariant and convex risk measure. For more details see Section 2.7 and Section 3.1.

$x \in K$, for every (nonempty) compact set $K \subseteq \mathbb{R}^n$. Therefore, assuming that $X$ is a nonempty compact set, it follows that w.p.1

$$\begin{aligned} \hat{v}_N^* &\to v^*, \text{ as } N \to \infty, && (1.0.58)\\ \mathbb{D}\left(\hat{S}_N, S\right) &\to 0, \text{ as } N \to \infty. && (1.0.59)\end{aligned}$$

This completes the presentation of the main results of [71].

Now we present some results concerning exponential rates of convergence for empirical estimates of some risk measures. It is worth mentioning that we are not aware of any publication obtaining large deviations bounds results for the empirical estimates of the optimal value and the optimal solutions of risk averse stochastic programming problems. Indeed, one of the contributions of our work is precisely to derive this type of results. There are, however, some large deviations results for empirical estimates of some important type of risk measures, such as the Average Value-at-Risk and more broadly the class of Optimized Certainty Equivalent (OCE) risk measures. In [12] the author derived an exponential rate of convergence for the empirical estimator of the $\mathrm{AV@R}_{1-\alpha}(\cdot)$ risk measure[27], for $\alpha \in [0, 1)$. He considered separately the upper and lower deviations of the statistical estimator with respect to its true counterpart. Different constants in the exponential rate of convergence were obtained for each side of the deviation, as well as different classes of risk measures were considered for each side. Let us present more details about these results. Let $Z$ be a bounded random variable that satisfies, without loss of generality, the following inequalities

$$0 \leq Z \leq U < +\infty \qquad (1.0.60)$$

w.p.1, where $U$ is a positive real number. Consider $N \in \mathbb{N}$ i.i.d. copies $\{Z^i : 1 \leq i \leq N\}$ of the random variable $Z$ defined on a common probability space. First we present the exponential bound for the upper deviation. In this setting, he obtained results for the class of OCE risk measures (see Section 3.1, in particular, Definition 3.1.1 and Remark 3.1.2). The OCE of a random variable $Z$ under $\phi \in \Phi$ (see Definition 3.1.1) is defined as

$$\mu_\phi(Z) := \inf_{s \in \mathbb{R}} \left\{ s + \mathbb{E}\phi\left(Z - s\right) \right\}. \qquad (1.0.61)$$

Its empirical estimator is given by

$$\hat{\mu}_\phi\left(Z^1, \ldots, Z^N\right) := \inf_{s \in \mathbb{R}} \left\{ s + \frac{1}{N} \sum_{i=1}^N \phi\left(Z^i - s\right) \right\}. \qquad (1.0.62)$$

[27]As it is common place in the literature the author denoted this risk measure by the term *Conditional Value-at-Risk*. We prefer to adopt this other nomenclature here in order to avoid any possible confusion with conditional risk measures considered in multistage risk averse stochastic programming problems (see Section 2.7).

The author proved that the following inequality holds[28]

$$\mathbb{P}\left[\hat{\mu}_\phi(Z^1,\ldots,Z^N) \geq \mu_\phi(Z)+\epsilon\right] \leq \exp\left\{-\frac{2\epsilon^2 N}{\phi(U)^2}\right\}, \tag{1.0.63}$$

for every $\epsilon > 0$ and $N \in \mathbb{N}$. The $\mathrm{AV@R}_{1-\alpha}(\cdot)$ risk measure, for $\alpha \in [0,1)$, is an example of OCE risk measure (see also Example 3.1.8). In fact, $\mathrm{AV@R}_{1-\alpha}(\cdot) = \mu_{\phi_{1-\alpha}}(\cdot)$, where $\phi_{1-\alpha}(z) := \frac{1}{1-\alpha}\max\{z,0\}$. In that case, equation (1.0.63) becomes

$$\mathbb{P}\left[\widehat{\mathrm{AV@R}}_{1-\alpha}(Z^1,\ldots,Z^N) \geq \mathrm{AV@R}_{1-\alpha}(Z)+\epsilon\right] \leq \exp\left\{-2\frac{(1-\alpha)^2\epsilon^2}{U^2}N\right\}. \tag{1.0.64}$$

For the lower deviation bound the author supposed additionally that $Z$ has a continuous distribution function. The lower deviation bound was obtained for the $\mathrm{AV@R}_{1-\alpha}$ risk measures, for $\alpha \in [0,1)$, rather than for the whole class of OCE risk measures (satisfying $\phi(U) < +\infty$). The author stated that

$$\mathbb{P}\left[\widehat{\mathrm{AV@R}}_{1-\alpha}(Z^1,\ldots,Z^N) \leq \mathrm{AV@R}_{1-\alpha}(Z)-\epsilon\right] \leq 3\exp\left\{-(1/5)(1-\alpha)\frac{\epsilon^2}{U^2}N\right\}, \tag{1.0.65}$$

for any $\epsilon > 0$ and $N \in \mathbb{N}$. Note that this bound is sharper than the other one for sufficiently large $N$ for $\alpha$ greater than 0.9[29]. It is worth mentioning that the upper deviation bound (1.0.63) was derived using: (a) the McDiarmid's bounded difference inequality (see [44]) to obtain an exponential bound for

$$\mathbb{P}\left[\hat{\mu}_\phi(Z^1,\ldots,Z^N) - \mathbb{E}\hat{\mu}_\phi(Z_1,\ldots,Z_N) \geq \epsilon\right], \tag{1.0.66}$$

and (b) the fact that $\mathbb{E}\hat{\mu}_\phi(Z_1,\ldots,Z_N) \leq \mu_\phi(Z)$ (see [12, Proposition 3.1]).

In [73, Section 6.6.1] the authors derived statistical properties of the empirical estimator of the Average Value-at-Risk risk measure. In their analysis it was taken as given an integrable (not necessarily bounded) random variable $Z$ and a sample $\{Z^1,\ldots,Z^N\}$ of i.i.d. random variables defined on the same probability space having the same distribution as $Z$. Besides proving the strong consistency of the empirical estimator and besides deriving asymptotic results for this estimator, the authors also derived large deviations-type bounds for the convergence of $\widehat{\mathrm{AV@R}}_{1-\alpha}(Z^1,\ldots,Z^N)$ to $\mathrm{AV@R}_{1-\alpha}(Z)$. They followed a different approach from the one used in [12]. It is worth mentioning that our results were derived following a similar approach, although there are important differences between their range of applicability and some involved hypotheses. We consider statistical estimators related to the optimal

[28]Since $\phi \in \Phi$ and $U > 0$, we have that $0 < \phi(U) \leq +\infty$. Note that the inequality (1.0.63) is trivially satisfied if $\phi(U) = +\infty$, however, in that case, we do not obtain an exponential rate of convergence. When $\phi(U) < +\infty$, it follows that $\beta = 2\left(\epsilon/\phi(U)\right)^2$ is positive.

[29]On risk management applications one usually takes $\alpha \in \{0.95, 0.99, 0.997\}$.

value and optimal solutions of a risk averse stochastic programming problem whereas in [73, Section 6.6.1] the result was obtained for a fixed random variable $Z$. Indeed, our results also extend to the multistage setting. Moreover, while in [73, Section 6.6.1] the results are obtained for the Average Value-at-Risk risk measure, our results are applicable to a broader class of risk measures. In fact, we derive large deviations-type results for the class of OCE risk measures, where $\phi \in \Phi$ is assumed Lipschitz continuous. For finishing this brief comparison it is worth mentioning that differently from [12] and similarly to [73, Section 6.6.1] we do not restrict our analysis to bounded random variables. Now we give more details about exponential bound results obtained in [73, Section 6.6.1].

Let us begin by recalling the fact that (see also Proposition 3.1.13):

$$\operatorname*{argmin}_{s\in\mathbb{R}} \left\{ h(s) := s + \frac{1}{1-\alpha}\mathbb{E}\left[Z-s\right]_+ \right\} \tag{1.0.67}$$

is equal to the set of $\alpha$-quantiles $[q_\alpha^-(Z), q_\alpha^+(Z)]$ of $Z$ (see Section 2.3). This set is a nonempty bounded closed interval, whenever $\alpha \in (0,1)$. So, let us suppose that $\alpha \in (0,1)$[30]. Taking any real numbers $u < q_\alpha^-(Z) \le q_\alpha^+(Z) < U$ , it follows that $\Delta := \frac{1}{2}\min\{h(u) - \mathrm{AV@R}_{1-\alpha}(Z), h(U) - \mathrm{AV@R}_{1-\alpha}(Z)\} > 0$. For $u \le s \le U$, assume that the random variables $W_s := [Z-s]_+ - \mathbb{E}\left[Z-s\right]_+$ are $\sigma$-sub-Gaussian[31], that is,

$$M_s(z) := \mathbb{E}\exp\left(W_s z\right) \le \exp\{\sigma^2 z^2/2\}, \text{ for all } z \in \mathbb{R}. \tag{1.0.68}$$

It can be shown that (see [73, Proposition 6.63]):

$$\mathbb{P}\left[\left|\widehat{\mathrm{AV@R}}_{1-\alpha}(Z^1,\ldots,Z^N) - \mathrm{AV@R}_{1-\alpha}(Z)\right| \ge \epsilon\right] \le \tfrac{8(U-u)}{(1-\alpha)\epsilon}\exp\left[-\tfrac{N\epsilon^2(1-\alpha)^2}{32\sigma^2}\right], \tag{1.0.69}$$

for any $0 < \epsilon < \Delta$. Therefore, given a confidence level of $\theta \in (0,1)$, if we take the sample size

$$N \ge \log\left(\frac{8(U-u)}{(1-\alpha)\theta\epsilon}\right)\frac{32\sigma^2}{(1-\alpha)^2\epsilon^2}, \tag{1.0.70}$$

then the probability in (1.0.69) is less than or equal to $\theta$. This completes the presentation of selected results in [73, Section 6.6.1].

As we have said previously, to the best of our knowledge sample complexity estimates for risk averse stochastic programming problems were not derived yet

[30] When $\alpha = 0$ it is a well-known fact (see also Proposition 3.1.11 and Remark 3.1.12) that $\mathrm{AV@R}_1(Z) = \mathbb{E}Z$, for any integrable random variable $Z$.

[31] In [73, Section 6.6.1] the authors suppose that the family of random variables $W_s$, for $u \le s \le U$, are $(\sigma, a)$-sub-exponential instead of $\sigma$-sub-Gaussian, where $\sigma$ and $a$ are positive constants. That means that the inequality (1.0.68) is satisfied for $|z| \le a$ rather than for all $z \in \mathbb{R}$. Of course, every $\sigma$-sub-Gaussian random variable is a $(\sigma, a)$-sub-exponential random variable, but the converse is not true in general. Therefore, we are assuming here a more strict condition. This simplifies a little bit the exposition.

in the literature. One contribution of this thesis is to derive sample complexity estimates for a class of risk averse stochastic programming problems. We derive results for *static* or *two-stage* problems and for *dynamic* or *multistage* problems that have a finite number of stages.

In this thesis, we consider the class of Optimized Certainty Equivalent (OCE) risk measures. As far as we can tell the class of OCE risk measures is a sufficiently broad class. In fact, many risk averse stochastic problems solved in practice [19, 39, 51, 57, 70, 72, 76] adopts a risk measure that belongs to this class[32] (see also Example 3.1.8). We recall some properties of OCE risk measures in Section 3.1 and establish new results that we use in the sequel.

The remainder of this thesis is organized as follows. In Chapter 2 we present some background material, such as known propositions and definitions, that are used along the thesis. Our objective in doing so is to make the thesis more self-contained. Although we prove some results in this chapter, in many occasions the proofs are omitted and we just give a pointer to the standard literature in case one wishes to have more details. In Sections 2.1.1 and 2.1.2 we give more details about the derivation of sample complexity estimates for static and dynamic, respectively, risk neutral stochastic programming problems. We derive the results by using the uniform exponential bound theorem (see Theorem 2.1.5). We follow this approach here because we use this particular tool for extending the sample complexity results to the risk averse setting. As mentioned previously, we extend in Section 2.1.2 the analysis done in [69]. Here we derive the results considering slightly weaker regularity assumptions. Moreover, we allow the parameters $T$ and $\delta$ to be, respectively, $T \geq 3$ and $0 \leq \delta < \epsilon$, instead of $T = 3$ and $\delta = \epsilon/2$. These are minor differences with respect to [68]. What we consider is the most important difference, is that working directly with $T$-stage problems it is possible to show that the sample complexity of multistage problems grows even faster than what a first look in the estimate provided in [69] for 3-stage problems might suggest. The remainder of Chapter 2 can be skipped without loss and it can be consulted as the need arises.

Starting from Chapter 3, most of the results presented in this thesis are new. In Chapter 3 we derive the sample complexity estimates for static stochastic programming problems with OCE risk measures. We recall in Section 3.1 the definition of this class of risk measures and provide some of its properties. In Section 3.2 we present the extended formulation for this class of problems, and explain the approach used for deriving the sample complexity estimates by using the theory already developed for risk neutral problems. It is worth mentioning that the developed theory cannot be applied directly, since when we deal with the extended formulation the feasible set of the optimization problem becomes unbounded. We

[32] Although it is worth noting that most of these references never mention that the considered risk measure is an OCE risk measure.

properly derive the sample complexity estimates in Section 3.3. We obtain a lower bound for the probability of the event (1.0.14) that is valid for every $N \in \mathbb{N}$ and that approaches one exponentially fast when we make $N$ to infinity. The estimate is akin to the one obtained for the risk neutral case, although the dependence on some problem data parameters differs. One important difference is the dependence with respect to the parameter $L(\phi)$, that is the Lipschitz constant of the function $\phi \in \Phi$ (see Section 3.1). Our estimate also provides a theoretical guarantee that *static* or *two-stage* stochastic programming problems with OCE risk measures can be efficiently approximated by Monte Carlo sampling-based approaches. Now, the result depends not only in not asking for obtaining a too much accurate solution of the true problem ($\epsilon > 0$ too small), but also in not using an OCE risk measure with $L(\phi)$ that is too large.

In Chapter 4 we extend the results of Chapter 3 considering the *dynamic* or *multistage* stochastic programming problems with nested OCE risk measures. Similar to the risk neutral setting, the derived sample complexity estimates for multistage problems present an order of growth that is even faster than the exponential with respect to the number of stages $T$ (see (4.0.75)). Indeed, a multiplicative term $(T-1)^{2(T-1)}$ appears when one obtain an estimate for the total number of scenarios in the scenario tree.

One could ask if the sample complexity estimates obtained for stochastic programming problems are in some sense tight. One possibility would be that the derived estimates were too gross and much bigger than the "best possible" estimate. Maybe much smaller sample sizes could be sufficient for guaranteeing that with a desirable level of probability the SAA optimal solutions would be approximate optimal solutions of the true problem. In particular, one could ask if the exponential growth of the sample complexity estimates for multistage problems with respect to the number of stages is really an unavoidable phenomenon or if maybe we just have not derived sufficiently tight estimates that do not present this behavior and, in some sense, do not suffer from the curse of dimensionality. In Chapter 5 we have shown that an order of growth exhibited by the sample complexity estimates derived in [69] (see also Section 2.1.2) is unavoidable for some problems. We construct a family of risk neutral $T$-stage problems whose members satisfy all the regularity conditions assumed in order to derive the sample complexity estimates for multistage problems and show that the number of scenarios needed for obtaining approximate optimal solutions of the true problem with high probability grows even faster than the exponential function with respect to $T$. This study was published in [53].

# CHAPTER 2

## Background material and preliminary results

## 2.1 Risk neutral stochastic programming problems

### 2.1.1 The static case

In this section we recall some sample complexity results on the SAA method for static risk neutral stochastic programming problems. To the best of our knowledge, these results were first developed in [75] for general stochastic programming problems. This type of results were obtained previously for specialized kind of problems such as problems with *discrete* decision variables (see [37]) and problems with a finite number of scenarios (see [74]). Here we follow closely reference [73], although we derive the results as a direct consequence of the uniform exponential bound theorem[1] (see [73, Theorem 7.75]).

We consider the general static risk neutral stochastic programming problem (SRN-SPP):

$$\min_{x \in X} \{f(x) := \mathbb{E}F(x, \xi)\}, \tag{2.1.1}$$

[1]As a consequence, some constants appearing in our estimates differs from their counterparts in [73]. Of course, this is a minor difference between both presentations.

where $\xi = (\xi_1, \ldots, \xi_d)$ is a random vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$; $x \in \mathbb{R}^n$ are the decision variables; $X \subseteq \mathbb{R}^n$ is the feasible set and $F : \mathbb{R}^n \times \mathbb{R}^d \to \overline{\mathbb{R}}$ is a measurable function.

As discussed in Chapter 1, one difficulty in solving problem (2.1.1) is that it is not possible, in general, to evaluate with accuracy the objective function $f(x)$ at $x \in \mathbb{R}^n$. In fact, we have that $f(x)$ is a $d$-dimensional integral. When it is possible to obtain a random sample from $\xi$, the SAA method can be used to circumvent this difficulty. In the SAA method, one takes a random sample $\{\xi^1, \ldots, \xi^N\}$ of $\xi$ and considers the SAA problem:

$$\min_{x \in X} \left\{ \hat{f}_N(x) := \frac{1}{N} \sum_{i=1}^{N} F(x, \xi^i) \right\}. \tag{2.1.2}$$

In the static case we usually assume that it is possible to evaluate $F(x, \xi)$, for $x \in \mathbb{R}^n$ and $\xi \in \operatorname{supp}(\xi)$. It turns out that given a sample realization of $\xi$ the SAA's objective function $\hat{f}_N$ can now be evaluated accurately. The SAA problem is usually easier to solve than the "true" problem, although one must keep in mind that the problem we are really trying to solve is problem (2.1.1).

In a nutshell, the sample complexity of the SAA method for SRN-SPP studies how large the sample size $N$ should be in order for (approximate) solutions of problem (2.1.2) be approximate solutions of problem (2.1.1) with high probability. In order to obtain results of this nature, one must assume some regularity conditions to be fulfilled by the problem instance. Let us recall some notation before proceeding.

We denote the optimal values of problems (2.1.1) and (2.1.2) by:

$$v^* := \inf_{x \in X} f(x), \text{ and} \tag{2.1.3}$$

$$\hat{v}_N^* := \inf_{x \in X} \hat{f}_N(x), \tag{2.1.4}$$

respectively. Given $\epsilon \geq 0$, we denote the set of $\epsilon$-optimal solutions of problems (2.1.1) and (2.1.2) by:

$$S^\epsilon := \{x \in X : f(x) \leq v^* + \epsilon\}, \tag{2.1.5}$$

$$\hat{S}_N^\epsilon := \{x \in X : \hat{f}_N(x) \leq \hat{v}_N^* + \epsilon\}, \tag{2.1.6}$$

respectively. When dealing with exact optimal solutions, i.e. $\epsilon = 0$, we drop the superscript and write $S$ and $\hat{S}_N$ instead of $S^\epsilon$ and $\hat{S}_N^\epsilon$, respectively.

Let us assume that the true optimization problem is solvable, i.e. $S \neq \emptyset$[2]. We assume that the optimizer aims at obtaining an $\epsilon$-solution of the true problem, where $\epsilon > 0$ is a given tolerance parameter. For such, he will solve the SAA problem

[2] In fact, this will follow from the regularity conditions used for deriving the sample complexity estimates.

obtaining a $\delta$-optimal solution of this problem. This approach is guaranteed to work whenever every $\delta$-optimal solution of the SAA problem is an $\epsilon$-solution of the true problem, assuming that there exists a $\delta$-optimal solution of the SAA problem[3]. So let us consider the favorable event:

$$\left[\hat{S}_N^\delta \subseteq S^\epsilon\right] \bigcap \left[\hat{S}_N^\delta \neq \emptyset\right]. \tag{2.1.7}$$

More rigorously, in the study of the sample complexity of the SAA method one determines how large $N$ should be in order to:

$$\mathbb{P}\left(\left[\hat{S}_N^\delta \subseteq S^\epsilon\right] \bigcap \left[\hat{S}_N^\delta \neq \emptyset\right]\right) \geq 1-\theta, \tag{2.1.8}$$

where $\epsilon > 0$, $0 \leq \delta < \epsilon$ and $\theta \in (0,1)$ are the sample complexity parameters.

One strategy for deriving this kind of result consists of bounding from below the probability of the event:

$$\sup_{x \in X} \left|\hat{f}_N(x) - f(x)\right| \geq \frac{\epsilon - \delta}{2}, \tag{2.1.9}$$

in terms of $N$ and $\epsilon - \delta > 0$. The uniform exponential bound theorem is a key tool for obtaining such an estimate. Before presenting this theorem, let us consider the following regularity conditions:

(A1) For every $x \in X$, $f(x) = \mathbb{E}F(x,\xi)$ is finite.

(A2) There exists $\sigma \in \mathbb{R}_+$ such that $F(x,\xi) - f(x)$ is a $\sigma$-sub-Gaussian random variable, for every $x \in X$, that is:

$$M_x(s) := \mathbb{E}\exp\{s\,(F(x,\xi) - f(x))\} \leq \exp\{\sigma^2 s^2/2\},\ \forall s \in \mathbb{R}.$$

(A3) There exists a measurable function $\chi : \text{supp}(\xi) \to \mathbb{R}_+$ whose moment generating function $M_\chi(s)$ is finite, for $s$ in a neighborhood of zero, such that

$$|F(x,\xi) - F(x',\xi)| \leq \chi(\xi)\,\|x - x'\|, \tag{2.1.10}$$

for all $x', x \in X$ and $\xi \in E \subseteq \text{supp}\{\xi\}$, where $\mathbb{P}[\xi \in E] = 1$.

(A4) $X \subseteq \mathbb{R}^n$ is a nonempty compact set with diameter $D$.

(A5) $\{\xi^i : i \in \mathbb{N}\}$ is an independent and identically distributed (i.i.d.) sequence of random vectors defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\xi^1 \overset{d}{\sim} \xi$.

[3]If $\hat{S}_N^\delta = \emptyset$, then $\hat{S}_N^\delta \subseteq S^\epsilon$ immediately. However this situation is not favorable, since the optimizer does not obtain an $\epsilon$-solution of the true problem.

**Remark 2.1.1.** *In principle, the function $F(x, \xi)$ restricted to the set $X \times \operatorname{supp}(\xi)$ can assume the values $\pm\infty$ (see Example 2.1.2). Let us point out that this is not a frivolous mathematical generality. Later we will show that two-stage stochastic programming problems can be cast as static problems. Moreover, for two-stage stochastic problems, let us recall that $F(x, \xi) = +\infty$, whenever the feasible set of the second stage problem, that depends on $x$ and on $\xi$, is empty.*

*Of course, the regularity conditions (A1)-(A4) impose some restrictions on the problems instances that we will consider here. Assumption (A1) guarantees that $F(x, \xi)$ is finite for almost every $\xi \in \operatorname{supp}(\xi)$, since the random variable $F(x, \xi)$ has finite expected value. This means that there exists a measurable set $E_x \subseteq \operatorname{supp}(\xi)$ such that $\mathbb{P}[\xi \in E_x] = 1$ and $F(x, \xi) \in \mathbb{R}$, for every $\xi \in E_x$. For two-stage stochastic programming problems this implies that the recourse is relatively complete. Note that (A1) does not imply, in general, that there exists a measurable set $E$ that has probability 1 such that $F(x, \xi)$ is finite, for every $\xi \in E$ and $x \in X$. However, assumption (A3) does imply that stronger condition. In fact, by assumption (A3), it follows that:*

$$0 \leq |F(x', \xi) - F(x, \xi)| \leq \chi(\xi) \|x' - x\| < +\infty, \ \forall x', x \in X, \ \forall \xi \in E. \tag{2.1.11}$$

*Therefore, it follows that $F(x, \xi)$ is finite, for every $x \in X$ and $\xi \in E$.* □

The following example is a modification of the one presented in [9, Page 109]. We show that even if the problem data satisfies the regularity (A1)-(A4), the cost function $F(x, \xi)$ can assume non-finite values, for some $x \in X$ and $\xi \in \operatorname{supp} \xi$.

**Example 2.1.2.** *Consider the feasible set $X = [0, 1] \subseteq \mathbb{R}$ and the random data $\xi$ having the following c.d.f.*

$$H_\xi(z) = \begin{cases} 0, & \text{if } z \leq 0 \\ \exp\left\{-\dfrac{1}{z^3}\right\}, & \text{if } z > 0 \end{cases}. \tag{2.1.12}$$

*Note that $\operatorname{supp} \xi = [0, \infty)$. Now, consider the following function:*

$$F(x, \xi) := \inf \{y \geq 0 : \xi y = 1 - x\}. \tag{2.1.13}$$

*Note that $(x, 0) \in X \times \operatorname{supp} \xi$, for every $0 \leq x \leq 1$. Moreover, observe that $F(x, 0) = \infty$, for every $x \in [0, 1)$. Therefore, it is not true that $F(x, \xi)$ is finite, for every $x \in X$ and $\xi \in \operatorname{supp} \xi$. However, it is worth noting that $\mathbb{P}[\xi = 0] = 0$ and $F(x, \xi) = (1 - x)/\xi$ is finite, for every $x \in X$, whenever $\xi > 0$. Of course, $\mathbb{P}[\xi > 0] = 1$. In fact, one can verify that the problem data satisfy all the regularity conditions (A1)-(A4)[4]. This fact is an elementary consequence of the following:*

[4]Recall that condition (A5) is about the sampling method and not about the problem data.

**Claim 2.1.3.** *The random variable $Z = 1/\xi$ is a $\psi_2$-random variable (see Section 2.4).*

*Before proving the claim, let us show that it implies the validity of the conditions (A1)-(A4). (A4) is valid since $X = [0, 1]$. Since $Z$ is a $\psi_2$-random variable, it follows that $Z - \mathbb{E}Z$ is a $\sigma$-sub-Gaussian random variable, for some $\sigma > 0$. Since $F(x, \xi) = (1-x)Z$ w.p.1, it follows that $f(x) := \mathbb{E}F(x, \xi)$ is finite, for every $x \in [0, 1]$ and*

$$F(x, \xi) - f(x) = (1 - x)(Z - \mathbb{E}Z) \tag{2.1.14}$$

*is a $\sigma$-sub-Gaussian random variable, for every $x \in [0, 1]$. Finally, note that*

$$|F(x, \xi) - F(x', \xi)| = \frac{1}{\xi}|x - x'|, \tag{2.1.15}$$

*for every $x, x' \in [0, 1]$ and $\xi > 0$. Since $Z$ is a $\psi_2$-random variable, we have in particular that $M_Z(t)$ is finite, for every $t \in \mathbb{R}$. Thus, condition (A4) is also satisfied.*

*Proof.* (of the Claim 2.1.3) For showing that $Z$ is a $\psi_2$-random variable we verify that its tails decay sufficiently fast to zero (see Proposition 2.4.2), i.e., that there exists a finite $K > 0$ such that

$$\mathbb{P}[|Z| \geq s] \leq \exp\left(1 - s^2/K^2\right), \forall s \geq 0. \tag{2.1.16}$$

Take any $s > 0$. Since $\xi \geq 0$ w.p.1, we have that

$$\begin{aligned}
\mathbb{P}[|Z| \geq s] &= \mathbb{P}[Z \geq s] \\
&= \mathbb{P}[1/\xi \geq s] \\
&= \mathbb{P}[\xi \leq 1/s] \\
&= \exp\left\{-s^3\right\} \\
&\leq \exp\left\{1 - s^2\right\}.
\end{aligned}$$

Thus, (2.1.16) is satisfied with $K = 1$. This completes the proof of the claim. □

In the next proposition we show that, as a consequence of the previous regularity conditions, $S \neq \emptyset$ and $\mathbb{P}[\hat{S}_N \neq \emptyset] = 1$.

**Proposition 2.1.4.** *Let $N \in \mathbb{N}$ be given. The following statements hold:*

*(a) If conditions (A1) and (A3) hold, then $f : X \to \mathbb{R}$ is Lipschitz continuous on $X$.*

*(b) Assuming additionally that condition (A4) holds, then $S \neq \emptyset$.*

*(c) If conditions (A1) and (A3) − (A5) hold, then $\mathbb{P}[\hat{S}_N \neq \emptyset] = 1$.*

*Proof.* Suppose that (A1) and (A3) hold. By (A3) there exists a measurable $E \subseteq \text{supp}(\xi)$ that has probability 1 such that inequality (2.1.10) holds, for every $x', x \in X$ and $\xi \in E$. Moreover, since $\chi(\xi)$ has finite moment generating function in a neighborhood of zero, Proposition 2.8.6 implies that:

$$\mathbb{E}|\chi(\xi)|^k < +\infty, \tag{2.1.17}$$

for all $k \in \mathbb{N}$. In particular, we have that $M := \mathbb{E}\chi(\xi) < +\infty$. Since (A1) is also satisfied, we conclude that:

$$\begin{aligned}
|f(x) - f(x')| &= |\mathbb{E}F(x,\xi) - \mathbb{E}F(x',\xi)| & (2.1.18)\\
&= |\mathbb{E}[F(x,\xi) - F(x',\xi)]| & (2.1.19)\\
&\leq \mathbb{E}\,|F(x,\xi) - F(x',\xi)| & (2.1.20)\\
&= \mathbb{E}\left[|F(x,\xi) - F(x',\xi)|\,\mathbb{1}_{\{\xi\in E\}}\right] & (2.1.21)\\
&\leq \mathbb{E}\left[\chi(\xi)\,\|x - x'\|\,\mathbb{1}_{\{\xi\in E\}}\right] & (2.1.22)\\
&= M\,\|x - x'\|\,, & (2.1.23)
\end{aligned}$$

for all $x, x' \in X$. Observe that the second equality is valid because $F(x,\xi)$ and $F(x',\xi)$ are both integrable by (A1)[5]. The third and fourth equalities hold because $\mathbb{P}\,[\xi \in E] = 1$. We obtain that $f$ is $M$-Lipschitz continuous on $X$.

Additionally, if we assume that condition (A4) holds, we obtain that problem (2.1.1) has an optimal solution, since $X$ is a nonempty compact set.

Finally, let us consider that condition (A5) is also satisfied and let $N \in \mathbb{N}$ be given. Since $\{\xi^i : i \in \mathbb{N}\}$ are identically distributed and $\xi^1 \overset{d}{\sim} \xi$, we have that:

$$\mathbb{P}[\xi^i \in E] = 1, \tag{2.1.24}$$

for all $i \in \mathbb{N}$. It follows that:

$$1 \geq \mathbb{P}\left[\bigcap_{i=1}^{N}[\xi^i \in E]\right] \geq \mathbb{P}\left[\bigcap_{i=1}^{+\infty}[\xi^i \in E]\right] = 1. \tag{2.1.25}$$

When $\xi^i \in E$, for all $i = 1, \ldots, N$, we have that:

$$\left|F(x,\xi^i) - F(x',\xi^i)\right| \leq \chi(\xi^i)\,\|x - x'\|\,, \tag{2.1.26}$$

for all $x, x' \in X$ and $0 \leq \chi(\xi^i) < +\infty$. So,

$$\left|\hat{f}_N(x) - \hat{f}_N(x')\right| \leq \frac{1}{N}\sum_{i=1}^{N}\chi(\xi^i)\,\|x - x'\|\,, \tag{2.1.27}$$

for all $x, x' \in X$ and $\frac{1}{N}\sum_{i=1}^{N}\chi(\xi^i) < +\infty$, i.e. $\hat{f}_N$ is Lipschitz continuous on $X$ w.p.1. Therefore, $\mathbb{P}[\hat{S}_N \neq \emptyset] = 1$. □

[5] It is not true, in general, that $\mathbb{E}[Z-Y] = \mathbb{E}Z - \mathbb{E}Y$. Consider arbitrary (finite-valued) random variables $Z = Y$ satisfying $\mathbb{E}Z = +\infty$. Although $Z - Y = 0$ and $\mathbb{E}[Z-Y] = 0$, we have that $\mathbb{E}Z - \mathbb{E}Y = +\infty - (+\infty) = +\infty$.

Now we state the uniform exponential bound theorem (see [73, Theorem 7.75]).

**Theorem 2.1.5.** *Consider a general SRN-SPP such as* (2.1.1) *and suppose that conditions (A1)-(A5) are satisfied. Take any* $\tilde{M} > M = \mathbb{E}\chi(\xi)$ *finite. Then, for* $\epsilon > 0$ *and* $N \in \mathbb{N}$*, we have that:*

$$\mathbb{P}\left[\sup_{x\in X}\left|\hat{f}_N(x) - f(x)\right| \geq \epsilon\right] \leq \exp\{-N\mathsf{m}\} + 2\left[\frac{2\rho D\tilde{M}}{\epsilon}\right]^n \exp\left\{-\frac{N\epsilon^2}{32\sigma^2}\right\}, \quad (2.1.28)$$

*where* $\mathsf{m} \in (0, +\infty]$ *is a quantity depending on* $\tilde{M}$ *and* $\chi(\xi)$*,* $D$ *and* $\sigma^2$ *are constants depending on the problem data, and* $\rho$ *is a universal constant.*

*Proof.* For a proof see [73, Sec. 7.2.10], particularly Theorems 7.73 and 7.75. Let us point out that assumptions (A1) and (A5) were stated in the first paragraph of reference's section (see page 450 of [73]). Moreover, conditions (C2) and (C3) of the reference were agglutinated in assumption (A2) in our presentation. Finally, condition (C4) is just assumption (A2). □

**Remark 2.1.6.** *Our notation is slightly different from that of reference [73]. Here we denote by* $M$ *the expected value of* $\chi(\xi)$*, instead of* $L$*. We proceed in that way in order to avoid any possible confusion to the constant* $L(\phi)$ *that will be defined when we present the OCE risk measures (see section 3.1). The constant* $\mathsf{m} \in (0, +\infty]$ *was denoted by* $l$ *on [73, Theorem 7.75]. Observe that it can assume the value* $+\infty$ *and, in that case, we assume the following convention:*

$$\exp\{-\infty\} := \lim_{x\to-\infty} \exp\{x\} = 0. \quad (2.1.29)$$

□

**Remark 2.1.7.** *(A glimpse on Large Deviation Theory) In order to give more details about how* $\mathsf{m}$ *was obtained from the problem data, we give a very short presentation of some concepts and result of the large deviation theory. All the results presented here are taken from [73, Section 7.2.9]. For a more detailed presentation of this topic, the reader should also consult [18, Chapter 2]. Let* $Z$ *be a random variable. We denote its moment generating function by* $M_Z(s) := \mathbb{E}\exp\{sZ\} \in (0, +\infty]$*, for every* $s \in \mathbb{R}$*. It is well-known that* $M_Z(\cdot)$ *is a convex function,* $M_Z(0) = 1$ *and* $\operatorname{dom} M_Z$ *is an interval of* $\mathbb{R}$*. Suppose that* $\mathbb{E}Z = \mu \in \mathbb{R}$*. Given* $\{Z_1, \ldots, Z_N\}$ *i.i.d. copies of* $Z$*, the upper bound of Cramer's Large Deviation theorem gives*

$$\mathbb{P}\left[\frac{1}{N}\sum_{i=1}^{N} Z_i \geq z\right] \leq \exp\{-NI_Z(z)\} \quad (2.1.30)$$

*for every* $z \geq \mu$*, where* $I_Z(\cdot) := (\log M_Z)^*(\cdot)$ *is known as the LD rate function of* $Z$ *and* $^*$ *denotes the convex conjugate operator. The rate function* $I_Z(\cdot)$ *is a nonnegative* *convex function and* $I_Z(\mu) = 0$. *Note that* $I_Z(\cdot)$ *attains its minimum value at* $z = \mu$. *If the moment generating function* $M_Z(s)$ *is finite for* $s$ *in a neighborhood of* $s = 0$, *then* $I_Z(z) \in (0, +\infty]$, *for any* $z \neq 0$, *and satisfies*

$$I_Z(z) = \frac{(z-\mu)^2}{2\sigma^2} + \mathrm{o}\left(|a-\mu|^2\right), \tag{2.1.31}$$

*where* $\sigma^2 := \mathrm{Var}\,[Z] < +\infty$. *This finishes our presentation of this topic.*

*Now, let us consider the definition of* $\mathsf{m}$. *Assuming that* $\chi(\xi)$ *has finite moment generating function in a neighborhood of* $0$ *and taking any* $\tilde{M} > M := \mathbb{E}\chi(\xi) \in \mathbb{R}$, $\mathsf{m}$ *is given by* $I_\chi\left(\tilde{M}\right) \in (0, +\infty]$. *If* $\{\xi^1, \ldots, \xi^N\}$ *are i.i.d. copies of* $\xi$, *then*

$$\mathbb{P}\left[\frac{1}{N}\sum_{i=1}^{N} \chi(\xi^i) \geq \tilde{M}\right] \leq \exp\left\{-N\mathsf{m}\right\} \tag{2.1.32}$$

*for every* $N \in \mathbb{N}$. *A similar bound is also valid for* $\mathbb{P}\left[\frac{1}{N}\sum_{i=1}^{N}\chi(\xi^i) \leq \tilde{M}\right]$ *where* $\tilde{M} < M$ *and* $\mathsf{m} = I_\chi(\tilde{M}) \in (0, +\infty]$. *This means that the sample mean* $\frac{1}{N}\sum_{i=1}^{N}\chi(\xi^i)$ *concentrates around the expected value* $M$ *exponentially fast with respect to the sample size* $N$. □

**Remark 2.1.8.** *We have shown in Proposition 2.8.3 that the absolute constant* $\rho > 0$ *appearing in equation (2.1.28) is less than or equal to* $5$. □

Theorem 2.1.5 shows that the SAA objective function $\hat{f}_N(x)$ converges in probability to $f(x)$, uniformly on $X$, as $N \to +\infty$. Now we present as its corollary the sample complexity estimate of the SAA method for SRN-SPP satisfying the stated regularity conditions.

**Corollary 2.1.9.** *Consider a general SRN-SPP such as* (2.1.1) *and suppose that conditions* (*A*1) − (*A*5) *are satisfied. Take any* $\tilde{M} > M = \mathbb{E}\chi(\xi)$. *Let* $\epsilon > 0$, $0 \leq \delta < \epsilon$ *and* $N \in \mathbb{N}$ *be given. We have that:*

$$\mathbb{P}\left(\left[\hat{S}_N^\delta \subseteq S^\epsilon\right] \cap [\hat{S}_N^\delta \neq \emptyset]\right) \geq 1 - \exp\{-N\mathsf{m}\} - 2\left[\frac{4\rho D\tilde{M}}{\epsilon-\delta}\right]^n \exp\left\{-\frac{N(\epsilon-\delta)^2}{128\sigma^2}\right\}, \tag{2.1.33}$$

*where* $\mathsf{m} := I_\chi(\tilde{M}) \in (0, +\infty]$, $D = \mathrm{diam}\, X$ *and* $\sigma^2$ *are constants depending on the problem data; and* $\rho$ *is a universal constant.*

*Proof.* We begin by showing that:

$$\left[\sup_{x\in X}\left|\hat{f}_N(x) - f(x)\right| \leq \frac{\epsilon-\delta}{2}\right] \subseteq \left[\hat{S}_N^\delta \subseteq S^\epsilon\right]. \tag{2.1.34}$$

If $\{\xi^1, \ldots, \xi^N\}$ is such that $\hat{S}_N^\delta = \emptyset$, then there is nothing to be done: $\hat{S}_N^\delta \subseteq S^\epsilon$ trivially. Suppose that $\{\xi^1, \ldots, \xi^N\}$ is such that the events $\left[\sup_{x\in X}\left|\hat{f}_N(x) - f(x)\right| \leq \frac{\epsilon-\delta}{2}\right]$ and $\left[\hat{S}_N^\delta \neq \emptyset\right]$ occur. Assume $x \in \hat{S}_N^\delta$. We have that:

$$
\begin{aligned}
f(x) - \frac{\epsilon-\delta}{2} &\leq \hat{f}_N(x) && (2.1.35)\\
&\leq \hat{v}_N^* + \delta && (2.1.36)\\
&\leq \left(v^* + \frac{\epsilon-\delta}{2}\right) + \delta && (2.1.37)\\
&= v^* + \frac{\epsilon+\delta}{2}, && (2.1.38)
\end{aligned}
$$

i.e. $f(x) \leq v^* + \epsilon$, therefore $x \in S^\epsilon$. Let us point out that we have applied Proposition 2.8.4 (observe that $\inf_{x\in X} f(x) > -\infty$ by Proposition 2.1.4) in the third inequality above in order to conclude that:

$$
\begin{aligned}
|\hat{v}_N^* - v^*| &= \left|\inf_{x\in X}\hat{f}_N(x) - \inf_{x\in X} f(x)\right| && (2.1.39)\\
&\leq \sup_{x\in X}\left|\hat{f}_N(x) - f(x)\right|. && (2.1.40)
\end{aligned}
$$

We have also used the fact that the event $\left[\sup_{x\in X}\left|\hat{f}_N(x) - f(x)\right| \leq \frac{\epsilon-\delta}{2}\right]$ occurs. By Proposition 2.1.4 we have that:

$$
1 \geq \mathbb{P}[\hat{S}_N^\delta \neq \emptyset] \geq \mathbb{P}[\hat{S}_N \neq \emptyset] = 1, \qquad (2.1.41)
$$

i.e. $\mathbb{P}[\hat{S}_N^\delta \neq \emptyset] = 1$. Applying Theorem 2.1.5 we conclude that:

$$
\begin{aligned}
\mathbb{P}\left(\left[\hat{S}_N^\delta \subseteq S^\epsilon\right] \cap [\hat{S}_N^\delta \neq \emptyset]\right) &= \mathbb{P}\left[\hat{S}_N^\delta \subseteq S^\epsilon\right] && (2.1.42)\\
&\geq \mathbb{P}\left[\sup_{x\in X}\left|\hat{f}_N(x) - f(x)\right| < \frac{\epsilon-\delta}{2}\right] && (2.1.43)\\
&\geq 1 - \exp\{-N\mathsf{m}\} && (2.1.44)\\
&\quad -2\left[\frac{4\rho D\tilde{M}}{\epsilon-\delta}\right]^n \exp\left\{-\frac{N(\epsilon-\delta)^2}{128\sigma^2}\right\}, && (2.1.45)
\end{aligned}
$$

which proves the corollary. □

Now, we present the result in terms of the three sample complexity parameters: $\epsilon > 0$, $0 \leq \delta < \epsilon$ and $\theta \in (0, 1)$. Given real numbers $a, b$, we denote $\max\{a, b\}$ by $a \vee b$.

**Corollary 2.1.10.** *Consider a general SRN-SPP such as* (2.1.1) *and suppose that conditions* (*A*1) − (*A*5) *are satisfied. Take any* $\tilde{M} > M$. *Let* $\epsilon > 0$, $0 \leq \delta < \epsilon$,

$\theta \in (0,1)$ *and* $N \in \mathbb{N}$ *be given. If the sample size* $N$ *satisfies:*

$$N \geq \frac{128\sigma^2}{(\epsilon-\delta)^2}\left[n\log\left(\frac{4\rho D\tilde{M}}{\epsilon-\delta}\right)+\log\left(\frac{4}{\theta}\right)\right]\bigvee\left[\frac{1}{\mathsf{m}}\log\left(\frac{2}{\theta}\right)\right], \tag{2.1.46}$$

*where the constants* $\mathsf{m}$, $M$ *and* $\rho$ *are as in the previous corollary, then:*

$$\mathbb{P}\left(\left[\hat{S}_N^{\delta} \subseteq S^{\epsilon}\right] \cap [\hat{S}_N^{\delta} \neq \emptyset]\right) \geq 1-\theta. \tag{2.1.47}$$

*Proof.* Given $\epsilon > 0$, $0 \leq \delta < \epsilon$ and $\theta \in (0,1)$, just take $N$ sufficiently large so that:

$$\exp\{-N\mathsf{m}\} \leq \frac{\theta}{2}, \text{ and} \tag{2.1.48}$$

$$2\left[\frac{4\rho D\tilde{M}}{\epsilon-\delta}\right]^n \exp\left\{-\frac{N(\epsilon-\delta)^2}{128\sigma^2}\right\} \leq \frac{\theta}{2} \tag{2.1.49}$$

are satisfied. The result follows from the previous corollary. □

Now, let us present an important class of problems, known as two-stage stochastic programming problems. As we have briefly discussed previously, this class of problems can be cast as static stochastic programming problems. Consider the general two-stage (risk neutral) stochastic programming problem:

$$\min_{x_1 \in X_1}\left\{F_1(x_1)+\mathbb{E}\left[\inf_{x_2 \in X_2(x_1,\xi)} F_2(x_2,\xi)\right]\right\}, \tag{2.1.50}$$

where $\xi = (\xi_1, \ldots, \xi_d)$ is the random data; $x_1 \in \mathbb{R}^{n_1}$ and $x_2 \in \mathbb{R}^{n_2}$ are the first and second stage decision variables, respectively; $F_1 : \mathbb{R}^{n_1} \to \mathbb{R}$ is a continuous function and $F_2 : \mathbb{R}^{n_2} \times \mathbb{R}^d \to \mathbb{R}$ is a normal integrand (see Definition 2.6.12); $X_1 \subseteq \mathbb{R}^{n_1}$ is the feasible set of the first stage problem and $X_2 : \mathbb{R}^{n_1} \times \mathbb{R}^d \rightrightarrows \mathbb{R}^{n_2}$ is a closed-valued multifunction (see Definition 2.6.2), where $X_2(x,\cdot)$ is measurable (see Definition 2.6.8), for every $x \in \mathbb{R}^{n_1}$.

The optimal value function of the second stage problem is known as the recourse function $Q_2 : \mathbb{R}^{n_1} \times \mathbb{R}^d \to \overline{\mathbb{R}}$ that is given by:

$$Q_2(x_1,\xi) := \inf_{x_2 \in X_2(x_1,\xi)} F_2(x_2,\xi). \tag{2.1.51}$$

Under the conditions considered above, $Q_2(x_1,\cdot)$ is a measurable function, for any $x_1 \in \mathbb{R}^{n_1}$. In fact, given $x_1 \in \mathbb{R}^{n_1}$, we have that $X_2(x_1,\cdot)$ is a closed-valued measurable multifunction and $F_2$ is a normal integrand. It follows from Corollary 2.6.17 that $Q_2(x_1,\cdot)$ is measurable.

In this class of problems, the optimizer has to make a decision $x_1$ in the first stage before knowing the realization of the random data $\xi$. After that, he observes

$\xi$ and must solve the second stage problem, given that $x_1$ was already chosen and that $\xi$ is known:

$$\min_{x_2 \in X_2(x_1,\xi)} F_2(x_2,\xi). \tag{2.1.52}$$

For casting two-stage stochastic programming problems as static ones, first note that we can write:

$$F(x_1,\xi) := F_1(x_1) + Q_2(x_1,\xi). \tag{2.1.53}$$

Moreover, note that the objective-function of problem (2.1.50) is:

$$F_1(x_1) + \mathbb{E}\left[Q_2(x_1,\xi)\right] = \mathbb{E}\left[F_1(x_1) + Q_2(x_1,\xi)\right] = \mathbb{E}F(x_1,\xi). \tag{2.1.54}$$

So, the same sample complexity results are valid for two-stage stochastic programming problems. For closing this section let us just rewrite the regularity conditions for two-stage problems. This will prepare the terrain to present the regularity conditions for multistage stochastic programming problems.

(A1') For every $x_1 \in X_1$, $\mathcal{Q}_2(x_1) := \mathbb{E}Q_2(x_1,\xi)$ is finite.

(A2') There exists $\sigma \in \mathbb{R}_+$ such that $Q_2(x_1,\xi) - \mathcal{Q}_2(x_1)$ is a $\sigma$-sub-Gaussian random variable, for all $x_1 \in X_1$, that is:

$$M_{x_1}(s) := \mathbb{E}\exp\{s\left(Q_2(x_1,\xi) - \mathcal{Q}_2(x_1)\right)\} \leq \exp\{\sigma^2 s^2/2\}, \ \forall s \in \mathbb{R}.$$

(A3') There exists a measurable function $\chi : \mathrm{supp}(\xi) \to \mathbb{R}_+$ whose moment generating function $M_\chi(s)$ is finite, for $s$ in a neighborhood of 0, such that:

$$\left|Q_2(x_1,\xi) - Q_2(x_1',\xi)\right| \leq \chi(\xi)\left\|x - x'\right\|,$$

for all $x_1', x_1 \in X$ and $\xi \in E \subseteq \mathrm{supp}\,\xi$, where $\mathbb{P}\left[\xi \in E\right] = 1$.

(A4') $X_1 \subseteq \mathbb{R}^{n_1}$ is a nonempty compact set with diameter $D_1$.

(A5') $\{\xi^i : i \in \mathbb{N}\}$ is an i.i.d. sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\xi^1 \overset{d}{\sim} \xi$.

**Remark 2.1.11.** *Writing $F(x_1,\xi) = F_1(x_1) + Q_2(x_1,\xi)$, for every $x_1 \in \mathbb{R}^{n_1}$ and $\xi \in \mathbb{R}^d$, and $X = X_1$, it follows that all the regularity conditions above are equivalent to the ones considered previously, excepting for condition (A3'). In fact,*

$$\left|F(x_1,\xi) - F(x_1',\xi)\right| = \left|F_1(x_1) + Q_2(x_1,\xi) - F_1(x_1') - Q_2(x_1',\xi)\right| \tag{2.1.55}$$

*that we cannot bound, in general, by an expression akin to the right hand side of (2.1.10). Indeed, $F_1$ was assumed just continuous, instead of Lipschitz continuous*[6]

[6]Consider, for instance, $Q_2(x_1,\xi) = Q_2(x_1',\xi)$, for all $x_1, x_1' \in X_1$ and $\xi \in \mathbb{R}^d$. Note that item (A3') is satisfied, although item (A3) is not, if we take a non-Lipschitz continuous function $F_1 : X_1 \to \mathbb{R}$.

*This difference is irrelevant in order to derive the sample complexity results for two-stage stochastic programming problems. In fact, note that:*

$$\begin{aligned} f(x_1) &= F_1(x_1) + \mathbb{E}Q_2(x_1,\xi), \text{ and} \quad (2.1.56)\\ \hat{f}_N(x_1) &= F_1(x_1) + \frac{1}{N}\sum_{i=1}^{N} Q_2(x_1,\xi^i). \quad (2.1.57) \end{aligned}$$

*So, the difference:*

$$\left|\hat{f}_N(x_1) - f(x_1)\right| = \left|\frac{1}{N}\sum_{i=1}^{N} Q_2(x_1,\xi^i) - \mathbb{E}Q_2(x_1,\xi)\right|$$

*does not depend on the function $F_1$. It follows that the uniform exponential bound theorem can be applied for two-stage stochastic programming problems assuming (A3') instead of (A3).* □

This finishes the presentation of the sample complexity results for static and two-stage risk neutral stochastic programming problems. In the next section we consider multistage problems.

### 2.1.2 The dynamic case

In this section we present the sample complexity estimates of the SAA method for multistage risk neutral stochastic programming problems. To the best of our knowledge, these results were first derived in [69].

Let us begin with the problem formulation. We follow closely [73, Section 3.1]. Consider the $T$-stage risk neutral stochastic programming problem

$$\begin{aligned} \min_{x_1\in X_1} \Bigg\{ F_1(x_1) + \mathbb{E}_{|\xi_1}\Bigg[ \inf_{x_2\in X_2(x_1,\xi_2)} F_2(x_2,\xi_2) \quad &+ \mathbb{E}_{|\xi_{[2]}}\Big[\ldots \\ + \mathbb{E}_{|\xi_{[T-1]}}\Bigg[\inf_{x_T\in X_T(x_{T-1},\xi_T)} F_T(x_T,\xi_T)\Bigg] &\ldots \Big]\Bigg]\Bigg\}, \end{aligned} \quad (2.1.58)$$

where $\{\xi_1,...,\xi_T\}$ is a stochastic process defined on some probability space $(\Omega,\mathcal{F},\mathbb{P})$. For every $t=1,\ldots,T$, $\xi_t$ is a $d_t$-dimensional random vector and $\xi_1$ is deterministic. Here, $x_t\in\mathbb{R}^{n_t}, t=1,...,T$, are the decisions variables, $F_t:\mathbb{R}^{n_t}\times\mathbb{R}^{d_t}\to\mathbb{R}$, $t=2,\ldots,T$, are Carathéodory functions (see Definition 2.6.14), and $X_t:\mathbb{R}^{n_{t-1}}\times\mathbb{R}^{d_t}\rightrightarrows\mathbb{R}^{n_t}$, $t=2,...,T$, are measurable multifunctions. We assume that the function $F_1:\mathbb{R}^{n_1}\to\mathbb{R}$ is continuous, and $X_1\subseteq\mathbb{R}^{n_1}$ is a nonempty closed set. Unless otherwise stated, all these features are automatically assumed when we consider multistage stochastic programs in this thesis.

Program (2.1.58) has a finite number of stages $T\geq 2$. The case $T=2$ was already considered in the previous section. So, here we assume that $T\geq 3$, although the derived results are also valid for $T = 2$. For each $t = 1, \ldots, T$, $\xi_{[t]} := (\xi_1, \ldots, \xi_t)$ represents the history of the stochastic process up to stage $t$. In this kind of problems, one has to make a sequence of decisions $x_1, \cdots, x_T$ while the information about the random process unfolds sequentially at the beginning of each stage. The vector $\xi_1$ is already known when the decision $x_1 \in X_1$ must be made, but the remaining random vectors $\xi_2, \ldots, \xi_T$ are unknown at this time. After $x_1$ is chosen, the decision maker observes a realization of the random vector $\xi_2$ and must choose $x_2 \in X_2(x_1, \xi_2)$ before knowing $\xi_3, \ldots, \xi_T$. The process continues stage after stage, and, at the final stage $t = T$, having already decided $x_{T-1} \in X_{T-1}(x_{T-2}, \xi_{T-1})$ in stage $T-1$ and having observed $\xi_T$ at the begin of stage $T$, the optimizer must choose $x_T \in X_T(x_{T-1}, \xi_T)$. Therefore, each decision $x_t$ should depend only on the history up to stage $t$, that is $\xi_{[t]}$, for $t = 1, \ldots, T$. These are commonly refereed in the stochastic programming literature as the nonanticipativity constraints.

The use of the conditional expectations in (2.1.58) instead of the unconditional expectations has to deal with the fact that when one must choose $x_t$ at the begin of stage $t$, the history of the process until stage $t$ is already known by the decision maker. In general, the distribution of the future uncertainties $\xi_{t+1}, \ldots, \xi_T$ could depend on $\xi_{[t]}$. Thus, one should make the decision $x_t$ in order to minimize the sum of the $t$-stage cost "$F_t(x_t, \xi_t)$" and the conditional expected costs of the future stages $t+1, \ldots, T$ given that $\xi_{[t]}$ happened.

In the nested formulation (2.1.58) one stresses the fact that the optimizer must solve a sequence of optimization problems at the begin of each stage $t = 1, \ldots, T$. This is particularly suitable for considering the dynamic programming equations[7]. Problem (2.1.58) can be written as

$$\min_{x_1 \in X_1} \left\{ f(x_1) := F_1(x_1) + \mathcal{Q}_2(x_1, \xi_1) \right\}, \tag{2.1.59}$$

where

$$\mathcal{Q}_2(x_1, \xi_1) := \mathbb{E}_{|\xi_1} \left[ Q_2(x_1, \xi_{[2]}) \right], \tag{2.1.60}$$

[7] An equivalent way to present multistage stochastic programming problems is to consider the decision variables $x_t$ as policy functions of the data process $\xi_{[t]}$ up to stage $t$. In this approach a solution candidate of problem (2.1.58) is a decision policy $\{x_t(\cdot) : t = 1, \ldots, T\}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ that is adapted to the filtration $\left\{\mathcal{F}_t := \sigma\left(\xi_{[t]}\right), t = 1, \ldots, T\right\}$ generated by the data process $\xi_{[t]}$ up to stage $t$ and that satisfies w.p.1 the feasibility conditions $x_1 \in X_1, x_t(\xi_{[t]}) \in X_t(x_{t-1}(\xi_{[t-1]}), \xi_t)$, $t = 2, \ldots, T$. A policy is said to be implementable if it is adapted to the filtration $\left\{\mathcal{F}_t := \sigma\left(\xi_{[t]}\right), t = 1, \ldots, T\right\}$ and it is said to be feasible if it satisfies w.p.1 the feasibility conditions $x_1 \in X_1, x_t(\xi_{[t]}) \in X_t(x_{t-1}(\xi_{[t-1]}), \xi_t)$, $t = 2, \ldots, T$. Therefore a solution (if it exists) of the $T$-stage stochastic programming problem is an implementable and feasible policy that minimizes the expected value $\mathbb{E}\left[F_1(x_1) + F_2(x_2(\xi_{[2]})) + \cdots + F_T(x_{[T]})\right]$. See Section [73, Section 3.1] for more information about this topic, in particular its Remark 3 where the equivalence between both formulations is discussed in details.

and

$$Q_2(x_1, \xi_{[2]}) = \inf_{x_2 \in X_2(x_1,\xi_2)} \Bigg\{ F_2(x_2, \xi_2) + \mathbb{E}_{\xi_{[2]}} \Bigg[ \inf_{x_3 \in X_3(x_2,\xi_3)} F_3(x_3, \xi_3) + \mathbb{E}_{\xi_{[3]}} [\dots \\ + \mathbb{E}_{|\xi_{[T-1]}} \Bigg[ \inf_{x_T \in X_T(x_{T-1},\xi_T)} F_T(x_T, \xi_T) \Bigg] \dots \Bigg] \Bigg\}, \tag{2.1.61}$$

Given the history process $\xi_{[2]}$ up to stage $t = 2$, the function $Q_2(x_1, \xi_{[2]})$ relates each first-stage decision $x_1$ to the conditional expected value of the total costs of the stages $t = 2, \dots, T$ assuming that the optimizer chooses the best possible decisions $x_2, \dots, x_T$ on the remaining stages $t = 2, \dots, T$. Of course when we say that decisions $x_2, \dots, x_T$ are the best possible ones, we are considering a specific criterion to compare future decisions. In risk neutral problems, the conditional expectation given the available information about the data process is the functional used to compare future decisions. Note also that $Q_2(x_1, \xi_{[2]})$ is $\mathcal{F}_2$-measurable, thus its value is unknown in stage 1. In stage 1, the optimizer must choose $x_1$ in order to minimize the sum of the first stage costs $F_1(x_1)$ and the expected recourse cost $\mathcal{Q}_2(x_1, \xi_1) = \mathbb{E}_{|\xi_1} Q_2(x_1, \xi_{[2]})$. The function $Q_2(x_1, \xi_{[2]})$ is known as a cost-to-go function and it can also be obtained recursively, going backward in stages. For seeing this, let us consider the last-stage problem

$$\min_{x_T \in X_T(x_{t-1},\xi_T)} F_T(x_T, \xi_T) \tag{2.1.62}$$

whose optimal value $Q_T(x_{T-1}, \xi_T)$ depends on the decision vector $x_{T-1}$ and data $\xi_T$. Although the decision $x_{T-1}$ depends on $\xi_{[T-1]}$, note that $Q_T : \mathbb{R}^{n_T} \times \mathbb{R}^{d_T} \to \overline{\mathbb{R}}$ does not depend directly on the terms $\xi_1, \dots, \xi_{T-1}$. Observe that the last term of the sum in (2.1.58) is

$$\mathcal{Q}_T(x_{T-1}, \xi_{[T-1]}) := \mathbb{E}_{|\xi_{[T-1]}} \left[ Q_T(x_{T-1}, \xi_T) \right]. \tag{2.1.63}$$

This function depends directly on the variables $\xi_{[T-1]}$, although $Q_T(x_{T-1}, \xi_T)$ does not. This dependence is due only to the conditional expectation in (2.1.63). Given $x_{T-2}$ and data $\xi_{T-1}$, the optimal value of the $(T-1)$-stage problem is

$$Q_{T-1}(x_{T-2}, \xi_{[T-1]}) = \inf_{x_{T-1} \in X_{T-1}(x_{T-2},\xi_{T-1})} \left\{ F_{T-1}(x_{T-1}, \xi_{T-1}) + \mathcal{Q}_T(x_{T-1}, \xi_{[T-1]}) \right\} \tag{2.1.64}$$

Again, we can consider the expected value of this random variable with respect to $\xi_{[T-2]}$

$$\mathcal{Q}_{T-1}(x_{T-2}, \xi_{[T-2]}) := \mathbb{E}_{|\xi_{[T-2]}} \left[ Q_{T-1}(x_{T-2}, \xi_{T-1}) \right]. \tag{2.1.65}$$

Continuing this process backward until $t = 2$, we obtain that

$$Q_2(x_1, \xi_{[2]}) = \inf_{x_2 \in X_2(x_1,\xi_2)} \left\{ F_2(x_2, \xi_2) + \mathcal{Q}_3(x_2, \xi_{[2]}) \right\} \tag{2.1.66}$$

obtaining the mentioned recursion relationship.

The optimal value of problem (2.1.59) is

$$v^* := \inf_{x_1 \in X_1} \left\{ f(x_1) = F_1(x_1) + \mathcal{Q}_2(x_1, \xi_1) \right\}, \tag{2.1.67}$$

and its set of $\epsilon$-solutions is given by

$$S^\epsilon := \left\{ x_1 \in X_1 : f(x_1) \leq v^* + \epsilon \right\}, \tag{2.1.68}$$

where $\epsilon \geq 0$. When $\epsilon = 0$ we write just $S$ instead of $S^0$. Note that $S$ is the set of first stage optimal solutions of the multistage problem considering the policy formulation. The policy and nested formulations are linked by the dynamic programming equations

$$Q_t(x_{t-1}, \xi_{[t]}) = \inf_{x_t \in X_t(x_{t-1}, \xi_t)} \left\{ F_t(x_t, \xi_t) + \mathcal{Q}_{t+1}(x_t, \xi_{[t]}) \right\}, \tag{2.1.69}$$

for $t = 2, \ldots, T$[8] in the following sense: an implementable policy $\bar{x}_t(\xi_{[t]})$, $t = 1, \ldots, T$, is optimal if and only if $\bar{x}_1$ is an optimal solution of the first stage problem, i.e. $\bar{x}_1 \in S$, and for $t = 2, \ldots, T$

$$\bar{x}_t(\xi_{[t]}) \in \operatorname*{argmin}_{x_t \in X_t(\bar{x}_{t-1}(\xi_{[t-1]}), \xi_t)} \left\{ F_t(x_t, \xi_t) + \mathcal{Q}_{t+1}(x_t, \xi_{[t]}) \right\} \text{ w.p.1.} \tag{2.1.70}$$

Note that in the policy formulation the optimal solution of a multistage stochastic programming problem depends on the process data, since $\bar{x}_t(\cdot)$ is $\mathcal{F}_t$-measurable. In the sequel we discuss how one can use Monte Carlo sampling-based methods to approximate problem (2.1.58) by one which has a finite number of scenarios. We continue to denote the approximating problem as the SAA problem. Note that a solution of the SAA problem considering the policy formulation consists of functions of the sampled scenarios that are adapted to the sample history process. Therefore, in general, the SAA optimal policy is not an implementable policy for the true problem. However, one can still consider how well the first stage solution of the SAA problem approximates the first stage solution of the true problem.

The sample complexity estimates in the multistage setting are obtained assuming that the random data $\xi_2, \ldots, \xi_T$ are stagewise independent. Let us present some of the consequences of this assumption. First, note that under the stagewise independent hypothesis the conditional distribution of $\xi_t$ given $\xi_{[t-1]}$ (or equivalently $\mathcal{F}_{t-1}$) is equal to its unconditional or marginal distribution, for every $t = 2, \ldots, T$. Moreover, for every Borel-measurable function $g : \mathbb{R}^{d_t} \to \mathbb{R}$ such that $g(\xi_t)$ is integrable,

$$\mathbb{E}_{|\xi_{[t-1]}} g(\xi_t) = \mathbb{E} g(\xi_t) \tag{2.1.71}$$

[8]For $t = T$, we define $\mathcal{Q}_{T+1}(\cdot)$ appearing in the right-side of (2.1.69) as the zero function.

is satisfied. Therefore, the following equality holds in the stagewise independent case

$$\begin{aligned} \mathcal{Q}_T(x_{T-1}, \xi_{[T-1]}) &:= \mathbb{E}_{|\xi_{[T-1]}}\left[Q_T(x_{T-1}, \xi_T)\right] \\ &= \mathbb{E}\left[Q_T(x_{T-1}, \xi_T)\right]. \end{aligned} \tag{2.1.72}$$

This means that $\mathcal{Q}_T(\cdot)$ does not depend on $\xi_{[T-1]}$, which implies that (see equation (2.1.63)) $Q_{T-1}(\cdot)$ does not depend on the entire history process $\xi_{[T-1]}$ up to stage $T-1$, but just on $\xi_{T-1}$ (and naturally also on $x_{T-2}$). Continuing backward in stages, we obtain that

$$Q_t(x_{t-1}, \xi_t) = \inf_{x_t \in X_t(x_{t-1}, \xi_t)} \left\{ F_t(x_t, \xi_t) + \mathcal{Q}_{t+1}(x_t) \right\}, \text{ and} \tag{2.1.73}$$

$$\mathcal{Q}_t(x_{t-1}) := \mathbb{E} Q_t(x_{t-1}, \xi_t), \tag{2.1.74}$$

for every $t = T-1, \ldots, 2$.

Let us recall that even in the static setting, it is in general not possible to evaluate accurately the objective function $f(x_1) := F_1(x_1) + \mathcal{Q}_2(x_1) = F_1(x_1) + \mathbb{E}Q_2(x_1, \xi_2)$ at a point $x_1 \in \mathbb{R}^{n_1}$. Naturally, the situation gets even worse in the multistage setting. Note that if $T \geq 3$, then even the accurate evaluation of $Q_2(x_1, \xi_2)$ is not possible anymore (except on trivial cases). Indeed, for evaluating $Q_2(x_1, \xi_2)$ exactly, one must solve a $(T-1)$-stage stochastic programming problem. Fortunately, akin to the static case, one can consider Monte Carlo sampling-based approaches for approximating problem (2.1.58) by a problem that has a finite (and hopefully manageable) number of scenarios.

We present two sampling schemes for obtaining a discrete state space stochastic process that approximates the original one. In both procedures we build a scenario tree to represent the stochastic process generated by the sampling scheme. We say that the generated stochastic process is the SAA stochastic process as opposed to the true stochastic process. In Section 2.2 we make an exposition about scenario trees. A scenario tree is composed by a set of nodes $\mathcal{N}$ and a set of arcs $\mathcal{A} \subseteq \mathcal{N} \times \mathcal{N}$. The nodes are organized in $T \geq 2$ levels, each one corresponding to the stages of the $T$-stage stochastic program. The set of nodes of level $t$ is denoted by $\mathcal{N}_t$. There is only one node at level 1 that is the root node $\iota_1$. Every node $\iota_{t+1}$ of level $t+1$ is connected by an arc $(\iota_t, \iota_{t+1}) \in \mathcal{A}$ to a unique node $\iota_t$ of level $t$, for $t = 1, \ldots, T-1$. In that case, we say that $\iota_t$ is the parent node $a(\iota_{t+1})$ of $\iota_{t+1}$. Equivalently, we say that $\iota_{t+1}$ is a child node of $\iota_t$. Of course, a node can have more than one child and we denote by $C_{\iota_t} \subseteq \mathcal{N}_{t+1}$ the set of all children nodes of $\iota_t$. The leaves nodes of the tree are the nodes at level $T$. These nodes do not have any children and are also denoted as the terminal nodes of the tree. In a scenario tree one moves from the root node at level $t = 1$ to a leaf node at level $t = T$ by following a path $(\iota_1, \iota_2, \ldots, \iota_T)$, where $(\iota_t, \iota_{t+1}) \in \mathcal{A}$, for $t = 1, \ldots, T-1$. A scenario tree is also composed by a family of positive real numbers $\rho = \{\rho_a : a \in \mathcal{A}\}$ satisfying condition (2.2.6). Take any $(\iota_t, \iota_{t+1}) \in \mathcal{A}$, where $\iota_t \in \mathcal{N}_t$, $\iota_{t+1} \in \mathcal{N}_{t+1}$ and $t < T$. The quantity $\rho_{(\iota_t, \iota_{t+1})}$ represents the conditional probability of moving from node $\iota_t$ to node $\iota_{t+1}$ given that we are at node $\iota_t$. Finally, a scenario tree is also composed by a family of real vectors $\xi = \{\xi_t^{\iota_t} \in \mathbb{R}^{n_t} : \iota_t \in \mathcal{N}_t, t = 1, \ldots, T\}$ that we assume satisfies condition (2.2.5). A key assumption for implementing the sampling schemes to be described in the sequel is that one is able to obtain independent observations $\xi_{t+1}^j$ of the random vector $\xi_{t+1}$ conditional to the history process $\xi_{[t]}$ up to stage $t < T$. So, the distribution of every $\xi_{t+1}^j$ is equal to the distribution of $\xi_{t+1}$ given $\xi_{[t]}$.

First we consider the *independent conditional sampling* scheme. We begin by associating the root node $\iota_1$ with the value of $\xi_1$ that is already known at stage 1, i.e. $\xi_1^{\iota_1} := \xi_1$. Then, we obtain an i.i.d. sample $\xi_2^j$, $j = 1, \ldots, N_2$, of $\xi_2$[9], and create as many nodes at level $t = 2$ as different values of $\{\xi_2^j : j = 1, \ldots, N_2\}$, connecting each created node $\iota_2$ to the root node $\iota_1$ through the arc $(\iota_1, \iota_2)$. By selecting appropriate indices $1 \leq j_1 < \cdots < j_{\operatorname{card} C_{\iota_1}} \leq N_2$, it follows that

$$\operatorname{card} C_{\iota_1} = \operatorname{card}\left\{\xi_2^j, j = 1, \ldots, N_2\right\} = \operatorname{card}\left\{\xi_2^{j_k}, k = 1, \ldots, \operatorname{card} C_{\iota_1}\right\} \leq N_2. \tag{2.1.75}$$

We associate to each node $\iota_2 \in C_{\iota_1}$ a vector $\xi_2^{\iota_2} := \xi_2^{j_k}$ in such a way that all different vectors are related to one node of $C_{\iota_1}$. For every $\iota_2 \in C_{\iota_1}$ define

$$\rho_{(\iota_1,\iota_2)} := \frac{\operatorname{card}\{1 \leq j \leq N_2 : \xi_2^j = \xi_2^{\iota_2}\}}{N_2} > 0. \tag{2.1.76}$$

This completes the definition of $\mathcal{N}_2 = C_{\iota_1}$. Note that $\sum_{\iota_2 \in C_{\iota_1}} \rho_{(\iota_1,\iota_2)} = 1$, and that if $\iota_2 \neq \iota_2'$, then $\xi_2^{\iota_2} \neq \xi_2^{\iota_2'}$.

We continue the definition of the scenario tree inductively for $2 \leq t < T$. Consider as given $\mathcal{N}_1 = \{\iota_1\}$, $\mathcal{N}_s = \cup_{\iota_{s-1} \in \mathcal{N}_{s-1}} C_{\iota_{s-1}}$, for $s = 2, \ldots, t$, and $\mathcal{A} = \{(\iota_s, \iota_{s+1}) : \iota_s \in \mathcal{N}_s, \iota_{s+1} \in C_{\iota_s}, s = 1, \ldots, t-1\}$. We also consider that there exists a family of vectors $\{\xi_s^{\iota_s} \in \mathbb{R}^{n_s} : \iota_s \in \mathcal{N}_s, s = 1, \ldots, t\}$ such that for every $2 \leq s \leq t$, if $\iota_s \neq \iota_s'$ are nodes at level $s$, there exists $2 \leq r \leq s$ such that

$$\xi_r^{a^{s-r}(\iota_s)} \neq \xi_r^{a^{s-r}(\iota_s')}. \tag{2.1.77}$$

Consider also as given a family of positive numbers $\{\rho_a : a \in \mathcal{A}\}$ satisfying

$$\sum_{\iota_{s+1} \in C_{\iota_s}} \rho_{(\iota_s,\iota_{s+1})} = 1, \tag{2.1.78}$$

for every $\iota_s \in \mathcal{N}_s$ and $s = 1, \ldots, t-1$. For every node $\iota_t \in \mathcal{N}_t$, we obtain a conditional i.i.d. sample $\{\xi_{t+1}^{\iota_t,j} : j = 1, \ldots, N_{t+1}\}$ of the random vector $\xi_{t+1}$ given that the history process up to stage $t$ is equal to $\xi_t = \xi_t^{\iota_t}$, $\xi_{t-1} = \xi_{t-1}^{a(\iota_t)}$, $\ldots$, $\xi_1 = \xi_1^{a^{t-1}(\iota_t)}$. We create as many children nodes $C_{\iota_t}$ as different values of $\{\xi_{t+1}^{\iota_t,j} : j = 1, \ldots, N_{t+1}\}$ and connect

[9] Since $\xi_1$ is deterministic, the marginal distribution of $\xi_2$ is equal to the conditional distribution of $\xi_2$ given $\xi_1$.

every created node $\iota_{t+1}$ with $\iota_t$ through the arc $(\iota_t, \iota_{t+1})$. By selecting appropriate indeces $1 \le j_1 < \cdots < j_{\mathrm{card}\, C_{\iota_t}} \le N_{t+1}$, it follows that

$$\mathrm{card}\, C_{\iota_t} = \mathrm{card}\left\{\xi_{t+1}^{\iota_t,j}, j=1,\ldots,N_{t+1}\right\} = \mathrm{card}\left\{\xi_{t+1}^{\iota_t,j_k}, k=1,\ldots,\mathrm{card}\, C_{\iota_t}\right\} \le N_{t+1}. \tag{2.1.79}$$

To every node $\iota_{t+1} \in C_{\iota_t}$, we associate a vector $\xi_{t+1}^{\iota_{t+1}} := \xi_{t+1}^{\iota_t,j_k}$ in such a way that all different vectors are related to one node of $C_{\iota_t}$. For every node $\iota_{t+1} \in C_{\iota_t}$, we define

$$\rho_{(\iota_t,\iota_{t+1})} := \frac{\mathrm{card}\{1 \le j \le N_{t+1} : \xi_{t+1}^j = \xi_{t+1}^{\iota_{t+1}}\}}{N_{t+1}} > 0. \tag{2.1.80}$$

This completes the definition of $C_{\iota_t}$. Repeating this procedure for every $\iota_t \in \mathcal{N}_t$, we construct all nodes $\mathcal{N}_{t+1}$ at level $t+1$

$$\mathcal{N}_{t+1} = \cup_{\iota_t \in \mathcal{N}_t} C_{\iota_t}. \tag{2.1.81}$$

Note that

$$\sum_{\iota_{t+1} \in C_{\iota_t}} \rho_{(\iota_t,\iota_{t+1})} = 1, \tag{2.1.82}$$

for every $\iota_t \in \mathcal{N}_t$. Let $\iota_{t+1} \neq \iota'_{t+1}$ be any nodes at level $t+1$. If $a(\iota_{t+1}) = a(\iota'_{t+1})$, then $\xi_{t+1}^{\iota_{t+1}} \neq \xi_{t+1}^{\iota'_{t+1}}$. Otherwise their parents nodes $\iota_t = a(\iota_{t+1})$ and $\iota'_t := a(\iota'_{t+1})$ are different, then by the induction hypothesis there exists $2 \le s \le t$,

$$\xi_s^{a^{t+1-s}(\iota_{t+1})} = \xi_s^{a^{t-s}(\iota_t)} \neq \xi_s^{a^{t-s}(\iota'_t)} = \xi_s^{a^{t+1-s}(\iota'_{t+1})}.$$

We assume that all conditional samples at each stage are independent of each other. This finishes the definition of the *independent conditional sampling* scheme.

This scheme can be used to approximate any stochastic process $\xi = (\xi_1, \ldots, \xi_T)$ by one that has finite state space. Note however that the scheme suffers from a serious drawback. To fix the ideas, suppose that for every $t = 2, \ldots, T$ and every realization $\xi_{[t-1]}$ of the history process up to stage $t-1$ the conditional distribution of the random vector $\xi_t$ given $\xi_{[t-1]}$ is continuous. Therefore, we draw conditional i.i.d. samples of a continuous random vector at every node of the tree. Moreover the samples are also independent from each other. This implies that w.p.1 all samples are different from each other. Thus, one needs to generate and store $\prod_{s=2}^t N_s$ vectors at each level $t = 2, \ldots, T$ of the tree. Assuming that $N_t \ge N \ge 2$, for every $t = 2, \ldots, T$, we have to generate and store at least

$$1 + N + \cdots + N^{T-1} = \frac{N^T - 1}{N - 1} = \mathrm{O}\left(N^{T-1}\right) \tag{2.1.83}$$

number of vectors. This number grows exponentially with respect to the number of stages $T$. This can be a serious limitation for the application of this scheme for multistage problems with a large number of stages.

In many applications one supposes that the stochastic process (or a suitable transformation of it) is stagewise independent. Even in that case, if one generates the scenario tree following the independent conditional sampling scheme, then, in general, the SAA stochastic process will not inherit the stagewise independence property from the true process. Many algorithms used to solve multistage stochastic programming problems, like the Stochastic Dual Dynamic Programming (SDDP) method (see [46]), take advantage of the stagewise independence hypothesis. In fact, we have already pointed that if the random data is stagewise independent, then the expected cost-to-go functions $\mathcal{Q}_t(x_t, \xi_{[t-1]}) = \mathcal{Q}_t(x_t)$, $t = 2, \ldots, T$, do not depend on the history process $\xi_{[t-1]}$ up to stage $t-1$. This property implies that the *cuts* generated in the stage $t$ of the backward step of the method improves all cost-to-go functions of this stage at once[10].

The previous remarks motivate the consideration of a sampling scheme that is able to approximate a stagewise independent stochastic process in such a way that:

(i) the number of samples does not grow exponentially with respect to the number of stages,

(ii) the constructed SAA stochastic process has stagewise indepedent random data.

The following scheme is known as the *identical conditional* sampling scheme and it accomplishes both tasks above. Given sample sizes $N_2, \ldots, N_T \in \mathbb{N}$, one draws i.i.d. samples $\xi_t^j$, $j = 1, \ldots, N_t$, of the marginal distribution of $\xi_t$, for $t = 1, \ldots, T$. We also assume that the set of random vectors

$$\left\{\xi_t^j : t = 2, \ldots, T,\ j = 1, \ldots, N_t\right\} \tag{2.1.84}$$

is independent.

The following steps of the procedure are similar. At level $t = 1$ of the tree we create the root node $\iota_1$ and relate with it the value of $\xi_1$, i.e. $\xi_1^{\iota_1} := \xi_1$. At level $t = 2$, we create as many nodes as different realizations of $\{\xi_t^j : j = 1, \ldots, N_2\}$, connecting each created node $\iota_2$ to the root node $\iota_1$ through the arc $(\iota_1, \iota_2)$. We select appropriate indeces $1 \leq j_1 < \cdots < j_{\operatorname{card} C_{\iota_1}}$ such that

$$\{\xi_2^j : j = 1, \ldots, N_2\} = \{\xi_2^{j_k} : k = 1, \ldots, \operatorname{card} C_{\iota_1}\}, \tag{2.1.85}$$

and associate to each node $\iota_2 \in C_{\iota_1}$ a vector $\xi_2^{\iota_2} := \xi_2^{j_k}$ in such a way that all different vectors are related to one node of $C_{\iota_1}$. For every $\iota_2 \in C_{\iota_1}$ define

$$\rho_{(\iota_1, \iota_2)} := \frac{\operatorname{card}\{1 \leq j \leq N_2 : \xi_2^j = \xi_2^{\iota_2}\}}{N_2} > 0. \tag{2.1.86}$$

[10]Saying in a different way, the generated cuts in the stage $t$ of the backward step are shared between all cost-to-go of stage $t$.

As in the previous sampling scheme, this completes the definition of $\mathcal{N}_2 = C_{\iota_1}$. We also have that $\sum_{\iota_2 \in C_{\iota_1}} \rho_{(\iota_1,\iota_2)} = 1$, and that if $\iota_2 \neq \iota_2'$, then $\xi_2^{\iota_2} \neq \xi_2^{\iota_2'}$.

For the remaining levels $t = 3, \ldots, T$ of the tree, we use the same sample $\xi_t^{\iota_{t-1},j} = \xi_t^j$, $j = 1, \ldots, N_t$, at every node $\iota_{t-1}$ at level $t-1$. We create as many children nodes for $\iota_{t-1}$ as different values of $\xi_t^j$, $j = 1, \ldots, N_t$, connecting every children node $\iota_t$ of $\iota_{t-1}$ through an arc $(\iota_{t-1}, \iota_t)$. We relate to each element of $C_{\iota_{t-1}}$ a different element of $\{\xi_t^j, j = 1, \ldots, N_t\}$ and define

$$\rho_{(\iota_{t-1},\iota_t)} := \frac{\operatorname{card}\{1 \leq j \leq N_t : \xi_t^j = \xi_t^{\iota_t}\}}{N_t} > 0, \tag{2.1.87}$$

for every $\iota_t \in C_{\iota_{t-1}}$. This finishes the definition of the scenario tree.

Given a scenario tree $(\mathcal{N}, \mathcal{A}, \{\xi_t^{\iota_t} : \iota_t \in \mathcal{N}_t, t = 1, \ldots, T\}, \{\rho_a : a \in \mathcal{A}\})$ constructed either way by some sampling scheme (such as any of the ones described previously), we consider its associated stochastic process[11] $\hat{\xi} = (\hat{\xi}_1, \hat{\xi}_2, \ldots, \hat{\xi}_T)$

$$\hat{\xi}(\iota_1, \ldots, \iota_T) := (\xi_1^{\iota_1}, \xi_2^{\iota_2}, \ldots, \xi_T^{\iota_T}) \tag{2.1.88}$$

defined on the set of all scenarios of the tree

$$\mathcal{S} = \{(\iota_1, \ldots, \iota_T) \in \mathcal{N}_1 \times \cdots \times \mathcal{N}_T : \iota_2 \in C_{\iota_1}, \iota_3 \in C_{\iota_2}, \ldots, \iota_T \in C_{\iota_{T-1}}\} \tag{2.1.89}$$

where each scenario $s = (\iota_1, \ldots, \iota_T) \in \mathcal{S}$ has the following probability of occurring

$$\rho_s = \prod_{t=1}^{T-1} \rho_{\iota_t, \iota_{t+1}}.$$

In the next proposition we show that if the scenario tree is constructed via the *identical conditional sampling* scheme, then $\hat{\xi}$ is stagewise independent. We also provide formulas in either sampling schemes for the conditional expectation of a function of $\hat{\xi}_{t+1}$ given the history process $\hat{\xi}_{[t]}$ up to stage $t$. This is particularly useful for deriving expressions of the cost-to-go functions under either the sampling schemes considered previously.

**Proposition 2.1.12.** *Take any integer $T \geq 3$ and let $N_2, \ldots, N_T \in \mathbb{N}$ be the sample sizes. The following assertions hold:*

(a) *If the scenario tree is constructed via the* identical conditional sampling *scheme, then $\hat{\xi} = \left(\hat{\xi}_1, \ldots, \hat{\xi}_T\right)$ is stagewise independent:*

$$\begin{aligned} \hat{\mathbb{P}}\left[\hat{\xi}_2 = \xi_2^{j_2}, \ldots, \hat{\xi}_T = \xi_T^{j_T}\right] &= \prod_{t=2}^{T} \frac{\operatorname{card}\{1 \leq i \leq N_t : \xi_t^i = \xi_t^{j_t}\}}{N_t} && (2.1.90) \\ &= \prod_{t=2}^{T} \hat{\mathbb{P}}\left[\hat{\xi}_t = \xi_t^{j_t}\right], && (2.1.91) \end{aligned}$$

*for $1 \leq j_t \leq N_t$ and $t = 2, \ldots, T$.*

[11]For more details, see Section 2.2.

(b) *Take any function* $g : \mathbb{R}^{t+1} \to \mathbb{R}$, *where* $1 \leq t \leq T-1$ *is given. If the scenario tree is constructed either via the* independent conditional sampling *scheme or the* identical conditional sampling *scheme, then*

$$\hat{\mathbb{E}}\left[ g\left(\hat{\xi}_{t+1}\right) \middle| \hat{\xi}_1 = \xi_1^{\iota_1}, \ldots, \hat{\xi}_t = \xi_t^{\iota_t} \right] = \frac{1}{N_{t+1}} \sum_{j=1}^{N_{t+1}} g\left(\xi_{t+1}^{\iota_t, j}\right). \tag{2.1.92}$$

**Remark 2.1.13.** *In item (b), equation (2.1.92) simplifies, as one should expect, to*

$$\hat{\mathbb{E}}\left[ g\left(\hat{\xi}_{t+1}\right) \middle| \hat{\xi}_1 = \xi_1^{\iota_1}, \ldots, \hat{\xi}_t = \xi_t^{\iota_t} \right] = \frac{1}{N_{t+1}} \sum_{j=1}^{N_{t+1}} g\left(\xi_{t+1}^{j}\right) = \hat{\mathbb{E}}\left[ g\left(\hat{\xi}_{t+1}\right)\right] \tag{2.1.93}$$

*when the scenario tree is constructed via the* identical conditional sampling *scheme.*
□

*Proof.* (a) Given $1 \leq j_t \leq N_t$, for $t = 2, \ldots, T$, there exists a unique scenario $(\iota_1, \iota_2^*, \ldots, \iota_T^*) \in \mathcal{S}$ such that $\xi_t^{\iota_t^*} = \xi_t^{j_t}$, for $t = 2, \ldots, T$. It follows that

$$\hat{\mathbb{P}}\left[\hat{\xi}_2 = \xi_2^{j_2}, \ldots, \hat{\xi}_T = \xi_T^{j_T}\right] = \rho_{(\iota_1, \iota_2^*, \ldots, \iota_T^*)} = \rho_{(\iota_1, \iota_2^*)} \rho_{(\iota_2^*, \iota_3^*)} \cdots \rho_{(\iota_{T-1}^*, \iota_T^*)} \tag{2.1.94}$$

Now, note that

$$\rho_{(\iota_{t-1}^*, \iota_t^*)} = \frac{\operatorname{card}\{1 \leq i \leq N_t : \xi_t^i = \xi_t^{\iota_t^*}\}}{N_t} \tag{2.1.95}$$

$$= \frac{\operatorname{card}\{1 \leq i \leq N_t : \xi_t^i = \xi_t^{j_t}\}}{N_t}, \tag{2.1.96}$$

since $\xi_t^{\iota_t^*} = \xi_t^{j_t}$, for $t = 2, \ldots, T$. This proves the first equality of item (a). Let us show the second one. Take any $1 \leq j_t \leq N_t$, where $2 \leq t \leq T$ is given. We have that

$$\hat{\mathbb{P}}\left[\hat{\xi}_t = \xi_t^{j_t}\right] = \sum_{\substack{(\iota_1, \ldots, \iota_T) \in \mathcal{S} \\ \xi_t^{\iota_t} = \xi_t^{j_t}}} \rho_{(\iota_1, \ldots, \iota_T)}. \tag{2.1.97}$$

Note that every node $\iota_{t-1}$ at level $t-1$ has one and only one child node $\iota_t(\iota_{t-1}) \in C_{\iota_{t-1}}$ satisfying $\xi_t^{\iota_t(\iota_{t-1})} = \xi_t^{j_t}$. Therefore, we can write the sum on the right-side of equation

(2.1.97) as the iterated sum

$$\begin{aligned}
\hat{\mathbb{P}}\left[\hat{\xi}_t = \xi_t^{j_t}\right] &= \sum_{\iota_2 \in C_{\iota_1}} \cdots \sum_{\iota_{t-1} \in C_{\iota_{t-2}}} \sum_{\iota_{t+1} \in C_{\iota_t(\iota_{t-1})}} \cdots \sum_{\iota_T \in C_{\iota_{T-1}}} \rho_{(\iota_1, \dots, \iota_{t-1}, \iota_t(\iota_{t-1}), \dots, \iota_T)} \\
&= \sum_{\iota_2 \in C_{\iota_1}} \rho_{(\iota_1, \iota_2)} \cdots \sum_{\iota_{t-1} \in C_{\iota_{t-2}}} \rho_{(\iota_{t-2}, \iota_{t-1})} \cdot \rho_{(\iota_{t-1}, \iota_t(\iota_{t-1}))} \cdots \sum_{\iota_T \in C_{\iota_{T-1}}} \rho_{(\iota_{T-1}, \iota_T)} \\
&= \sum_{\iota_2 \in C_{\iota_1}} \rho_{(\iota_1, \iota_2)} \cdots \sum_{\iota_{t-1} \in C_{\iota_{t-2}}} \rho_{(\iota_{t-2}, \iota_{t-1})} \cdot \rho_{(\iota_{t-1}, \iota_t(\iota_{t-1}))} \cdots \sum_{\iota_{T-1} \in C_{\iota_{T-2}}} \rho_{(\iota_{T-2}, \iota_{T-1})} \\
&= \qquad \vdots \\
&= \sum_{\iota_2 \in C_{\iota_1}} \rho_{(\iota_1, \iota_2)} \cdots \sum_{\iota_{t-1} \in C_{\iota_{t-2}}} \rho_{(\iota_{t-2}, \iota_{t-1})} \cdot \rho_{(\iota_{t-1}, \iota_t(\iota_{t-1}))},
\end{aligned}$$

where we have used above that $\sum_{\iota_s \in C_{\iota_{s-1}}} \rho_{(\iota_{s-1}, \iota_s)} = 1$, for every $\iota_{s-1} \in \mathcal{N}_{s-1}$ and $s = T, \dots, t+1$. Note also that

$$\begin{aligned}
\rho_{(\iota_{t-1}, \iota_t(\iota_{t-1}))} &= \frac{\operatorname{card}\{1 \le i \le N_t : \xi_t^i = \xi_t^{\iota_t(\iota_{t-1})}\}}{N_t} \qquad (2.1.98) \\
&= \frac{\operatorname{card}\{1 \le i \le N_t : \xi_t^i = \xi_t^{j_t}\}}{N_t}, \qquad (2.1.99)
\end{aligned}$$

for every $\iota_{t-1} \in \mathcal{N}_{t-1}$. Therefore, $\rho_{(\iota_{t-1}, \iota_t(\iota_{t-1}))} = C$ does not depend on $\iota_{t-1} \in \mathcal{N}_{t-1}$. It follows that

$$\begin{aligned}
\hat{\mathbb{P}}\left[\hat{\xi}_t = \xi_t^{j_t}\right] &= C \sum_{\iota_2 \in C_{\iota_1}} \rho_{(\iota_1, \iota_2)} \cdots \sum_{\iota_{t-1} \in C_{\iota_{t-2}}} \rho_{(\iota_{t-2}, \iota_{t-1})} \\
&= C \\
&= \frac{\operatorname{card}\{1 \le i \le N_t : \xi_t^i = \xi_t^{j_t}\}}{N_t},
\end{aligned}$$

using the fact that $\sum_{\iota_s \in C_{\iota_{s-1}}} \rho_{(\iota_{s-1}, \iota_s)} = 1$, for every $\iota_{s-1} \in \mathcal{N}_{s-1}$ and $s = 2, \dots, t-1$ in the second equality. This completes the proof of item (a).

(b) First, let us consider that the scenario tree was generated using the *independent conditional sampling* scheme. We have that

$$\hat{\mathbb{E}}\left[g\left(\hat{\xi}_{t+1}\right) \middle| \hat{\xi}_1 = \xi_1^{\iota_1}, \dots, \hat{\xi}_t = \xi_t^{\iota_t}\right] = \sum_{\iota_{t+1} \in C_{\iota_t}} \rho_{(\iota_t, \iota_{t+1})} g\left(\xi_{t+1}^{\iota_{t+1}}\right). \qquad (2.1.100)$$

For every $\iota_{t+1} \in C_{\iota_t}$, define $J(\iota_{t+1}) := \{1 \le j \le N_{t+1} : \xi_{t+1}^{\iota_t, j} = \xi_{t+1}^{\iota_{t+1}}\}$. We have that

$$\{1, \dots, N_{t+1}\} = \bigcup_{\iota_{t+1} \in C_{\iota_t}} J(\iota_{t+1}), \qquad (2.1.101)$$

where this union is disjoint and each one of its members is a nonempty set. Moreover,

$$\begin{aligned} \rho(\iota_t, \iota_{t+1}) g\left(\xi_{t+1}^{\iota_{t+1}}\right) &= \frac{1}{N_{t+1}} \operatorname{card} J(\iota_{t+1}) g\left(\xi_{t+1}^{\iota_{t+1}}\right) && (2.1.102)\\ &= \frac{1}{N_{t+1}} \sum_{j \in J(\iota_{t+1})} g\left(\xi_{t+1}^{\iota_{t+1}}\right) && (2.1.103)\\ &= \frac{1}{N_{t+1}} \sum_{j \in J(\iota_{t+1})} g\left(\xi_{t+1}^{\iota_t, j}\right) && (2.1.104) \end{aligned}$$

Summing up, we obtain

$$\begin{aligned} \hat{\mathbb{E}}\left[g\left(\hat{\xi}_{t+1}\right) \middle| \hat{\xi}_1 = \xi_1^{\iota_1}, \ldots, \hat{\xi}_t = \xi_t^{\iota_t}\right] &= \frac{1}{N_{t+1}} \sum_{\iota_{t+1} \in C_{\iota_t}} \sum_{j \in J(\iota_{t+1})} g\left(\xi_{t+1}^{\iota_t, j}\right) && (2.1.105)\\ &= \frac{1}{N_{t+1}} \sum_{j \in \bigcup_{\iota_{t+1} \in C_\iota} J(\iota_{t+1})} g\left(\xi_{t+1}^{\iota_t, j}\right) && (2.1.106)\\ &= \frac{1}{N_{t+1}} \sum_{j=1}^{N_{t+1}} g\left(\xi_{t+1}^{\iota_t, j}\right), && (2.1.107) \end{aligned}$$

which proves the result for the *independent conditional sampling* scheme. The result for the *identical conditional sampling* scheme follows immediately noting that $\xi_{t+1}^{\iota_t, j} = \xi_{t+1}^{j}$, for every $\iota_t \in \mathcal{N}_t$ and $j = 1, \ldots, N_{t+1}$, that is, we use the same sample $\{\xi_{t+1}^j : j = 1, \ldots, N_{t+1}\}$ for generating the children nodes of every node at level $t = 1, \ldots, T-1$. □

It is possible to obtain sample complexity estimates in the multistage setting for the SAA method considering that the scenario tree is generated using the *independent conditional sampling* scheme (see [69]). However, here we just present the results assuming that the scenario tree is generated via the *identical conditional sampling* scheme. As pointed out previously, this sampling scheme is particularly appealing for computational implementations. The following result is an immediate corollary of Proposition 2.1.12.

**Corollary 2.1.14.** *Take any integer $T \geq 3$ and let $N_2, \ldots, N_T \in \mathbb{N}$ be the sample sizes. If we generate the scenario tree using the* identical conditional sampling *scheme, then the following formulas for the SAA cost-to-go functions and the SAA expected cost-to-go functions hold:*

$$\begin{aligned} \hat{Q}_t\left(x_{t-1}, \xi_t^j\right) &= \inf_{x_t \in X_t\left(x_{t-1}, \xi_t^j\right)} \left\{F_t\left(x_t, \xi_t^j\right) + \hat{\mathcal{Q}}_{t+1}(x_t)\right\} && (2.1.108)\\ \hat{\mathcal{Q}}_t(x_{t-1}) &= \frac{1}{N_t} \sum_{j=1}^{N_t} \hat{Q}_t\left(x_{t-1}, \xi_t^j\right), && (2.1.109) \end{aligned}$$

*for* $1 \leq j \leq N_t$ *and* $t = 2, \ldots, T$[12].

*Proof.* The result follows immediately from item (a) of Proposition 2.1.12, since $\hat{\xi} = \left(\hat{\xi}_1, \ldots, \hat{\xi}_T\right)$ is stagewise independent. □

Before proceeding, we introduce some notation that will be used when we present in the sequel the regularity conditions for a $T$-stage stochastic programming problem:

$$\begin{aligned} \mathcal{X}_0 &:= \{0\} \subseteq \mathbb{R}, && (2.1.110)\\ X_1(x_0, \xi_1) &:= X_1,\ \forall x_0 \in \mathcal{X}_0, && (2.1.111)\\ \mathcal{Q}_{T+1}(x_T) &:= 0,\ \forall x_T \in \mathbb{R}^{n_{T+1}}. && (2.1.112) \end{aligned}$$

We introduce these objects only for the uniformity of notation. Now we enumerate some regularity conditions for a $T$-stage stochastic programming problem:

(M0) the random data $\xi_1, \xi_2, \ldots, \xi_T$ is stagewise independent.

(M1) the family of random vectors $\{\xi_t^j : j \in \mathbb{N}, t = 2, \ldots, T\}$ is independent and satisfies $\xi_t^j \overset{d}{\sim} \xi_t$, for all $j \in \mathbb{N}$, and $t = 2, \ldots, T$[13].

For each $t = 1, \ldots, T-1$:

(Mt.1) There exist a compact set $\mathcal{X}_t$ with finite diameter $D_t$ such that $X_t(x_{t-1}, \xi_t) \subseteq \mathcal{X}_t$, for every $x_{t-1} \in \mathcal{X}_{t-1}$ and $\xi_t \in \text{supp}(\xi_t)$.

(Mt.2) For every $x_t \in \mathcal{X}_t$, $\mathcal{Q}_{t+1}(x_t) = \mathbb{E}Q_{t+1}(x_t, \xi_{t+1})$ is finite.

(Mt.3) There exists a finite constant $\sigma_t > 0$ such that for any $x \in \mathcal{X}_t$, the following inequality holds

$$M_{t,x}(s) := \mathbb{E}\left[\exp\left(s(Q_{t+1}(x, \xi_{t+1}) - \mathcal{Q}_{t+1}(x)\right)\right] \leq \exp\left(\sigma_t^2 s^2/2\right),\ \forall s \in \mathbb{R}. \qquad (2.1.113)$$

(Mt.4) There exists a measurable function $\chi_t : \text{supp}(\xi_{t+1}) \to \mathbb{R}_+$ whose moment generating function $M_{\chi_t}(s)$ is finite, for $s$ in a neighborhood of zero, such that

$$|Q_{t+1}(x_t', \xi_{t+1}) - Q_{t+1}(x_t, \xi_{t+1})| \leq \chi_t(\xi_{t+1}) \|x_t' - x_t\| \qquad (2.1.114)$$

holds, for all $x_t', x_t \in \mathcal{X}_t$ and $\xi_{t+1} \in E_{t+1} \subseteq \text{supp}\, \xi_{t+1}$, where $\mathbb{P}[\xi_{t+1} \in E_{t+1}] = 1$.

(Mt.5) W.p.1 $\xi_{t+1}$ the multifunction $X_{t+1}(\cdot, \xi_{t+1})$ restricted to $\mathcal{X}_t$ is continuous (see Definition 2.6.3).

[12] As usual we consider the boundary condition $\hat{\mathcal{Q}}_{T+1}(x_T) = 0$, for every $x_T \in \mathbb{R}^{n_T}$.

[13] Since $\xi_1$ is deterministic, it is not necessary to sample from it.

We make some remarks about the stated conditions. Similar to the analysis done in [69] and [73, Section 5.8.2], here we are assuming the stagewise independence hypothesis (M0). As discussed previously this allows us to generate the scenario tree using the *identical conditional sampling* scheme. In this thesis we present the sample complexity estimates assuming always that the SAA scenario tree is constructed via the *identical conditional sampling* scheme. This is accomplished by considering the samples $\{\xi_t^j : 1 =, \ldots, N_t\}$, for $t = 2, \ldots, T$, where $N_2, \ldots, N_T$ are the sample sizes. (M1) asserts that this family of random vectors is independent and that $\xi_t^j \overset{d}{\sim} \xi_t$, for every $j = 1, \ldots, N_t$ and $t = 2, \ldots, T$ (compare with assumption (A5') of the static case). Item (M1.1) asserts that there exists a compact set $\mathcal{X}_1$ that contains $X_1$. Since $X_1$ is assumed nonempty closed (see the paragraph following equation (2.1.58)), this is equivalent to assume that $X_1$ is compact (compare with assumption (A4') in the static case). For the remainder of the stages $t = 2, \ldots, T-1$, it is assumed that the multifunctions $X_t(x_{t-1}, \xi_t)$ are uniformly bounded for $x_{t-1} \in \mathcal{X}_{t-1}$ and $\xi_t \in \operatorname{supp} \xi_t$. This allow us to apply the uniform exponential bound theorem considering appropriate family of random variables indexed in $\mathcal{X}_t$, for $t = 1, \ldots, T-1$. The set of assumptions (Mt.2) implies that the multistage problem has relatively complete recourse (compare with assumption (A1') of the static case). Assumptions (Mt.2) and (Mt.3) are akin to assumptions (A2') and (A3'), respectively, of the static case. Finally, assumptions (Mt.5) are technical conditions that guarantee that w.p.1. the SAA objective function $\hat{f}_{N_2,\ldots,N_T}(\cdot)$ is continuous on $\mathcal{X}_1$ (see Proposition 2.1.15). This implies that w.p.1 $\hat{S}_{N_2,\ldots,N_T} \neq \emptyset$.

Before proceeding let us introduce some notation. Whenever we assume conditions $(Mt.4)$, for $t = 1, \ldots, T-1$, we denote the expected value of $\chi_t(\xi_{t+1})$ by

$$M_t := \mathbb{E}\chi_t(\xi_{t+1}). \tag{2.1.115}$$

Since the random variable $\chi_t(\xi_{t+1}) \geq 0$ has finite moment generating function in a neighborhood of 0, it follows that $0 \leq M_t < +\infty$, for $t = 1, \ldots, T-1$. In Proposition 2.1.15 we consider some consequences of the stated regularity conditions.

**Proposition 2.1.15.** *Consider a general $T$-stage stochastic programming problem such as (2.1.58), where $T \geq 3$ is an arbitrary integer. The following assertions hold:*

*(a) If the problem satisfies the regularity conditions (M0), (Mt.1), (Mt.2), and (Mt.4), for $t = 1, \ldots, T-1$, then $Q_{t+1}(\cdot, \xi_{t+1})$ is a Lipschitz continuous function on $\mathcal{X}_t$ w.p.1, for $t = 1, \ldots, T-1$. It also follows that $\mathcal{X}_t \subseteq \operatorname{dom} X_{t+1}(\cdot, \xi_{t+1})$ w.p.1 and $\mathcal{Q}_{t+1}(\cdot)$ is $M_t$-Lipschitz continuous on $\mathcal{X}_t$, for $t = 1, \ldots, T-1$. In particular, we conclude that the first stage objective function*

$$f(x_1) = F_1(x_1) + \mathcal{Q}_2(x_1)$$

*of the true problem restricted to $x_1 \in \mathcal{X}_1$ is finite-valued and continuous and the set of first stage optimal solutions $S$ is nonempty.*

(b) *Consider given the sample sizes* $N_2, \ldots, N_T \in \mathbb{N}$. *If the problem satisfies the regularity conditions (M0), (M1), (Mt.1), (Mt.4) and (Mt.5), for* $t = 1, \ldots, T-1$, *and the SAA scenario tree is constructed using the* identical conditional sampling *scheme, then the SAA objective function* $\hat{f}_{N_2,\ldots,N_T}(x_1)$ *restricted to the set* $\mathcal{X}_1$ *is finite-valued and continuous w.p.1. In particular,* $\mathbb{P}\left[\hat{S}_{N_2,\ldots,N_T} \neq \emptyset\right] = 1$.

*Proof.* Let us recall that the stagewise independent hypothesis (M0) implies that the dependence of the cost-to-go functions $Q_{t+1}(\cdot,\cdot)$ relative to the history process $\xi_{[t+1]}$ up to stage $t+1$ is limited only to the random vector $\xi_{t+1}$ instead of $\xi_{[t+1]}$, for $t = 1, \ldots, T-1$. In particular, the expected cost-to-go functions $\mathcal{Q}_{t+1}(\cdot)$ do not depend on the random data. The proof of item (a) is similar to the proof of items (a) and (b) of Proposition 2.1.4. Note that condition (Mt.4) implies that, for every $\xi_{t+1} \in E_{t+1}$, $Q_{t+1}(\cdot, \xi_{t+1})$ is $\chi_t(\xi_{t+1})$-Lipschitz continuous on $\mathcal{X}_t$, where $E_{t+1} \subseteq \operatorname{supp} \xi_{t+1}$ satisfies

$$\mathbb{P}\left[\xi_{t+1} \in E_{t+1}\right] = 1,$$

for $t = 1, \ldots, T-1$. In particular, we have that $Q_{t+1}(\cdot, \xi_{t+1})$ is a finite-valued function on $\mathcal{X}_t$, for every $\xi_{t+1} \in E_{t+1}$. Since

$$Q_{t+1}(x_t, \xi_{t+1}) = \inf_{x_{t+1} \in X_{t+1}(x_t, \xi_{t+1})} \left\{F_{t+1}(x_{t+1}, \xi_{t+1}) + \mathcal{Q}_{t+2}(x_{t+1})\right\}, \tag{2.1.116}$$

it follows that $X_{t+1}(x_t, \xi_{t+1}) \neq \emptyset$, for all $x_t \in \mathcal{X}_t$ and $\xi_{t+1} \in E_{t+1}$, i.e. $\operatorname{dom} X_{t+1}(\cdot, \xi_{t+1}) \supseteq \mathcal{X}_t$, for all $t = 1, \ldots, T-1$. Assuming conditions (Mt.2), for $t = 1, \ldots, T-1$, it follows that

$$\begin{aligned} \left|\mathcal{Q}_{t+1}(x_t') - \mathcal{Q}_{t+1}(x_t)\right| &\leq \mathbb{E}\chi_t(\xi_{t+1}) \left\|x_t' - x_t\right\| && (2.1.117) \\ &= M_t \left\|x_t' - x_t\right\|, && (2.1.118) \end{aligned}$$

for every $x_t', x_t \in \mathcal{X}_t$. The first stage objective function of the true problem is just $f(x_1) = F_1(x_1) + \mathcal{Q}_2(x_1)$, where $F_1 : \mathbb{R}^n \to \mathbb{R}$ is a finite-valued continuous function. It follows that the restriction of $f$ to the compact set $\mathcal{X}_1 \supseteq X_1$ is continuous and that $S = \operatorname{argmin}_{x_1 \in X_1} f(x_1) \neq \emptyset$, since $X_1$ is nonempty compact. This completes the proof of item (a).

Now, we prove item (b). Condition (Mt.5) says that there exists $F_{t+1} \in \operatorname{supp} \xi_{t+1}$ satisfying $\mathbb{P}\left[\xi_{t+1} \in F_{t+1}\right] = 1$ such that $X_{t+1}(\cdot, \xi_{t+1}) : \mathcal{X}_t \rightrightarrows \mathcal{X}_{t+1}$ is a continuous multifunction, for every $\xi_{t+1} \in F_{t+1}$ and for every $t = 1, \ldots, T-1$. Since conditions (M1) and (Mt.4) also hold true, for $t = 1, \ldots, T-1$, we claim that the following event has probability 1

$$\mathcal{E} := \bigcap_{t=1}^{T-1} \bigcap_{j \in \mathbb{N}} \left[\xi_{t+1}^j \in E_{t+1} \cap F_{t+1}\right]. \tag{2.1.119}$$

This is a consequence of the following facts: (a) a countable intersection of almost sure events[14] is also an almost sure event, (b) $\mathbb{P}\left[\xi_{t+1} \in E_{t+1} \cap F_{t+1}\right] = 1$, for every $t = 1, \ldots, T-1$, and (c) $\xi_{t+1}^j \overset{d}{\sim} \xi_{t+1}$, for every $j \in \mathbb{N}$ and $t = 1, \ldots, T-1$[15].

Take any sample sizes $N_2, \ldots, N_T \in \mathbb{N}$. We have that

$$\mathcal{E}_{N_2,\ldots,N_T} := \bigcap_{t=1}^{T-1} \bigcap_{j=1}^{N_{t+1}} \left[\xi_{t+1}^j \in E_{t+1} \cap F_{t+1}\right] \supseteq \mathcal{E}, \tag{2.1.120}$$

therefore $\mathbb{P}\left(\mathcal{E}_{N_2,\ldots,N_T}\right) = 1$. Now, we show that whenever the event $\mathcal{E}_{N_2,\ldots,N_T}$ happens, every SAA expected cost-to-go function

$$\hat{\mathcal{Q}}_{t+1}(x_t) = \sum_{j=1}^{N_{t+1}} \hat{Q}_{t+1}(x_t, \xi_{t+1}^j), \ \forall x_t \in \mathcal{X}_t, \tag{2.1.121}$$

is finite-valued and continuous on $\mathcal{X}_t$, for $t = 1, \ldots, T-1$, where

$$\hat{Q}_{t+1}(x_t, \xi_{t+1}^j) = \inf_{x_{t+1} \in X_{t+1}(x_t, \xi_{t+1}^j)} \left\{F_{t+1}(x_{t+1}, \xi_{t+1}) + \hat{\mathcal{Q}}_{t+2}(x_{t+1})\right\} \tag{2.1.122}$$

are the SAA cost-to-go functions, for $t = 1, \ldots, T-1$. As usual we set $\hat{\mathcal{Q}}_{T+1}(x_T) := 0$, for every $x_T \in \mathcal{X}_T$, for uniformity of notation. For proving the result we show that if $\hat{\mathcal{Q}}_{t+1} : \mathcal{X}_t \to \mathbb{R}$ is finite-valued and continuous, then $\hat{\mathcal{Q}}_t : \mathcal{X}_{t-1} \to \mathbb{R}$ is also finite-valued and continuous, for $t = T, \ldots, 2$. We also verify that $\hat{\mathcal{Q}}_T(\cdot) : \mathcal{X}_{T-1} \to \mathbb{R}$ is finite-valued and continuous (base case in order to apply the induction step). Note that

$$\hat{Q}_T(x_{T-1}, \xi_T) = Q_T(x_{T-1}, \xi_T), \tag{2.1.123}$$

for every $x_{T-1} \in \mathcal{X}_{T-1}$ and $\xi_T \in \operatorname{supp} \xi_T$.

Whenever the event $\mathcal{E}_{N_2,\ldots,N_T}$ happens, we have in particular that $\xi_T^j \in E_T$, for every $j = 1, \ldots, N_T$. So, it follows from item (a) that

$$x_{T-1} \in \mathcal{X}_{T-1} \mapsto Q_T(x_{T-1}, \xi_T^j) \tag{2.1.124}$$

is finite-valued and continuous (since this is a Lipschitz continuous mapping), for every $j = 1, \ldots, N_T$. It follows that

$$\hat{\mathcal{Q}}_T(x_{T-1}) = \frac{1}{N_T} \sum_{j=1}^{N_T} Q_T(x_{T-1}, \xi_T^j), \ \forall x_{T-1} \in \mathcal{X}_{T-1}, \tag{2.1.125}$$

is finite-valued and continuous. This shows the base case.

[14] That is, events having probability 1.

[15] Note that we are not using here the fact that the random vectors $\{\xi_{t+1}^j : j \in \mathbb{N}, t = 1, \ldots, T-1\}$ are independent.

Now we prove the induction step. Assume that $\hat{\mathcal{Q}}_{t+1} : \mathcal{X}_t \to \mathbb{R}$ is finite-valued and continuous for some $t+1 \leq T$[16]. We claim that

$$\hat{Q}_t(x_{t-1}, \xi_t^j) = \inf_{x_t \in X_t(x_{t-1}, \xi_t^j)} \left\{ F_t(x_t, \xi_t^j) + \hat{\mathcal{Q}}_{t+1}(x_t) \right\}, \forall x_{t-1} \in \mathcal{X}_{t-1}, \tag{2.1.126}$$

is finite-valued and continuous, for every $j = 1, \ldots, N_t$, whenever the event $\mathcal{E}_{N_2,\ldots,N_T}$ happens. In fact, let us verify that the hypotheses of the Berge's Maximum Theorem (BMT) (see Proposition 2.6.4) hold whenever $\mathcal{E}_{N_2,\ldots,N_T}$ occurs. For such, take any $x_{t-1} \in \mathcal{X}_{t-1}$ and $1 \leq j \leq N_t$. Note that

(i) $x_{t-1} \in \mathcal{X}_{t-1} \subseteq \operatorname{dom} X_t(\cdot, \xi_t^j)$, since $\xi_t^j \in E_t$,

(ii) $g : \mathcal{X}_t \times \mathcal{X}_{t-1} \to \mathbb{R}$ defined as $g(x_t, x_{t-1}) := F_t(x_t, \xi_t^j) + \hat{\mathcal{Q}}_{t+1}(x_t)$ is continuous, since $F_t(\cdot, \cdot)$ is a Carathéodory function and $\hat{\mathcal{Q}}_{t+1}(\cdot)$ is continuous (induction hypothesis),

(iii) $X_t(\cdot, \xi_t^j)$ is continuous at $x_{t-1}$, since $\xi_t^j \in F_t$,

(iv) Defining $V := \mathcal{X}_{t-1}$ and noting that $X_t(V, \xi_t^j) \subseteq \mathcal{X}_t$ (condition (Mt.1)), it follows that $V$ is a neighborhood of $x_{t-1}$ in $\mathcal{X}_{t-1}$ and the image of $V$ through the continuous multifunction $X_t(\cdot, \xi_t^j)$ is a compact metric space (see Proposition 2.6.6).

Therefore, the BMT implies that $\hat{Q}_t(\cdot, \xi_t^j) : \mathcal{X}_{t-1} \to \mathbb{R}$ is continuous at $x_{t-1} \in \mathcal{X}_{t-1}$. Since $x_{t-1}$ and $1 \leq j \leq N_t$ are arbitrary, we conclude that $\hat{\mathcal{Q}}_t(\cdot) : \mathcal{X}_{t-1} \to \mathbb{R}$ is continuous. Thus, $\hat{f}_{N_2,\ldots,N_T} : \mathcal{X}_1 \to \mathbb{R}$ is continuous, whenever the event $\mathcal{E}_{N_2,\ldots,N_T}$ occurs. Finally, since $\mathcal{X}_1$ is compact, it follows that

$$\mathcal{E}_{N_2,\ldots,N_T} \subseteq \left[ \hat{S}_{N_2,\ldots,N_T} \neq \emptyset \right],$$

i.e. $\mathbb{P}\left[ \hat{S}_{N_2,\ldots,N_T} \neq \emptyset \right] = 1$. □

In Proposition 2.1.15 we considered sufficient conditions that guarantee the solvability of the true and the SAA stochastic programming problems[17]. Now we apply Theorem 2.1.5 for proving that, under appropriate regularity conditions, the probability that the first stage SAA objective function stays arbitrarily close to the first stage true objective function uniformly in $X_1$ approaches 1 exponentially fast with respect to the sample sizes $N_2, \ldots, N_T$. This result is an extension of Theorem 2.1.5 to the multistage setting. In the sequel, we apply it for obtaining the sample complexity estimates of the SAA method for risk neutral stochastic programming problems in the multistage setting.

[16] So, we are considering the range $t = 2, \ldots, T-1$.
[17] Of course, we can only state that the SAA problem is solvable w.p.1.

**Proposition 2.1.16.** *Take any integer $T \geq 3$. Consider a $T$-stage stochastic programming problem satisfying conditions (M0), (M1), and (Mt.1)-(Mt.4), for $t = 1, \ldots, T-1$. Denote the stage sample sizes by $N_2, \ldots, N_T \in \mathbb{N}$, and suppose that the scenario tree is constructed via the* identical conditional sampling *scheme. Let $\tilde{M}_t > M_t = \mathbb{E}\left[\chi_t(\xi_{t+1})\right] \in \mathbb{R}_+$ be given real numbers, $t = 1, \ldots, T-1$. Then, for any $\epsilon > 0$, the following estimate holds*

$$\mathbb{P}\left[\sup_{x_1 \in X_1} \left|\hat{f}_{N_2,\ldots,N_T}(x_1) - f(x_1)\right| \geq \epsilon\right] \leq \sum_{t=1}^{T-1} \Bigg( \exp\left\{-N_{t+1} I_{\chi_t}(\tilde{M}_t)\right\} + 2\left[\frac{2\rho D_t \tilde{M}_t}{\epsilon/(T-1)}\right]^{n_t} \exp\left\{-\frac{N_{t+1}\epsilon^2}{32\sigma_t^2 (T-1)^2}\right\} \Bigg), \tag{2.1.127}$$

*where $I_{\chi_t}(\cdot)$ is the LD rate function (see Remark 2.1.7) of the random variable $\chi_t(\xi_{t+1})$, for $t = 1, \ldots, T-1$.*

*Proof.* The idea of the proof is to bound from above w.p.1 the random quantity

$$\sup_{x_1 \in X_1} \left|\hat{f}_{N_2,\ldots,N_T}(x_1) - f(x_1)\right| \tag{2.1.128}$$

by a sum of random variables, such that we have a control of the tail decay of its terms with respect to the sample sizes $N_{t+1}$, for $t = 1, \ldots, T-1$. In fact, we show that we can take each term of the sum as

$$Z_t := \sup_{x_t \in X_t} \left| \frac{1}{N_{t+1}} \sum_{j=1}^{N_{t+1}} \left[Q_{t+1}(x_t, \xi_{t+1}^j) - \mathcal{Q}_{t+1}(x_t)\right] \right|, \; t = 1, \ldots, T-1. \tag{2.1.129}$$

In the final step, we apply Theorem 2.1.5 for each $Z_t$, $t = 1, \ldots, T-1$, obtaining an upper bound for the probability of $Z_t$ be greater or equal than $\epsilon/(T-1)$ as a function that depends on the problem data and on the sample size $N_{t+1}$.

From (M1) and (Mt.4), $t = 1, \ldots, T-1$, it follows that the event

$$\mathcal{E}_{N_2,\ldots,N_T} := \bigcap_{t=2}^{T} \bigcap_{j=1}^{N_t} \left[\xi_t^j \in E_t\right] \tag{2.1.130}$$

has probability 1, where $E_t$ are the measurable sets appearing in condition (Mt.4), for $t = 1, \ldots, T-1$. We claim that whenever the event $\mathcal{E}_{N_2,\ldots,N_T}$ happens,

$$\sup_{x_1 \in X_1} \left|\hat{f}_{N_2,\ldots,N_T}(x_1) - f(x_1)\right| \leq \sum_{t=1}^{T-1} Z_t. \tag{2.1.131}$$

Since $f(x_1) = F_1(x_1) + \mathcal{Q}_2(x_1)$, $\hat{f}_{N_2,\ldots,N_T}(x_1) = F_1(x_1) + \hat{\mathcal{Q}}_2(x_1)$ and $F_1(x_1)$ is finite, for every $x_1 \in X_1$, it follows that

$$\left|\hat{f}_{N_2,\ldots,N_T}(x_1) - f(x_1)\right| = \left|\hat{\mathcal{Q}}_2(x_1) - \mathcal{Q}_2(x_1)\right|, \tag{2.1.132}$$

for every $x_1 \in X_1$. Therefore, it is sufficient to bound from above the expression

$$\sup_{x_1 \in X_1} \left| \hat{\mathcal{Q}}_2(x_1) - \mathcal{Q}_2(x_1) \right|. \tag{2.1.133}$$

We divide the proof into two steps. In the first one, we show that whenever the event $\mathcal{E}_{N_2,\ldots,N_T}$ occurs the following inequality holds

$$\sup_{x_t \in \mathcal{X}_t} \left| \hat{\mathcal{Q}}_{t+1}(x_t) - \mathcal{Q}_{t+1}(x_t) \right| \leq Z_t + \sup_{x_{t+1} \in \mathcal{X}_{t+1}} \left| \hat{\mathcal{Q}}_{t+2}(x_{t+1}) - \mathcal{Q}_{t+2}(x_{t+1}) \right|, \tag{2.1.134}$$

for $t = 1, \ldots, T-1$. Let us prove this statement. Take any $x_t \in \mathcal{X}_t$, where $1 \leq t \leq T-1$ is arbitrary. By the triangular inequality, it follows that

$$\begin{aligned} \left| \hat{\mathcal{Q}}_{t+1}(x_t) - \mathcal{Q}_{t+1}(x_t) \right| \leq \; & \left| \tfrac{1}{N_{t+1}} \sum_{t=1}^{N_{t+1}} \left[ Q_{t+1}(x_t, \xi_t^j) - \mathcal{Q}_{t+1}(x_t) \right] \right| + \\ & \tfrac{1}{N_{t+1}} \sum_{t=1}^{N_{t+1}} \left| Q_{t+1}(x_t, \xi_t^j) - \hat{Q}_{t+1}(x_t, \xi_t^j) \right|, \end{aligned} \tag{2.1.135}$$

using also that

$$\hat{\mathcal{Q}}_{t+1}(x_t) = \frac{1}{N_{t+1}} \sum_{j=1}^{N_{t+1}} \hat{Q}_{t+1}(x_t, \xi_{t+1}^j). \tag{2.1.136}$$

The first term on the right-side of (2.1.135) is less than or equal to $Z_t$. Applying the inf-sup inequality (see Proposition 2.8.4) we obtain an upper bound for the second term. For this, we need to verify that: (i) $X_{t+1}(x_t, \xi_{t+1}^j) \neq \emptyset$, for every $x_t \in \mathcal{X}_t$ and for every $j = 1, \ldots, N_{t+1}$, and

$$(ii)\; Q_{t+1}\left(x_t, \xi_{t+1}^j\right) = \inf_{x_{t+1} \in X_{t+1}(x_t, \xi_{t+1}^j)} \left\{ F_{t+1}\left(x_{t+1}, \xi_{t+1}^j\right) + \mathcal{Q}_{t+2}(x_{t+1}) \right\} > -\infty,$$

for every $x_t \in \mathcal{X}_t$ and $j = 1, \ldots, N_{t+1}$. By item (a) of Proposition 2.1.15 both conditions (i) and (ii) hold whenever the event $\mathcal{E}_{N_2,\ldots,N_T}$ occurs. Therefore,

$$\begin{aligned} & \left| Q_{t+1}(x_t, \xi_{t+1}^j) - \hat{Q}_{t+1}(x_t, \xi_{t+1}^j) \right| = \\ & \left| \inf_{x \in X_{t+1}\left(x_t, \xi_{t+1}^j\right)} \left\{ F_{t+1}\left(x, \xi_{t+1}^j\right) + \mathcal{Q}_{t+2}(x) \right\} - \inf_{x \in X_{t+1}\left(x_t, \xi_{t+1}^j\right)} \left\{ F_{t+1}\left(x, \xi_{t+1}^j\right) + \hat{\mathcal{Q}}_{t+2}(x) \right\} \right| \\ & \qquad \leq \sup_{x \in X_{t+1}\left(x_t, \xi_{t+1}^j\right)} \left| \mathcal{Q}_{t+2}(x) - \hat{\mathcal{Q}}_{t+2}(x) \right| \\ & \qquad \leq \sup_{x \in \mathcal{X}_{t+1}} \left| \mathcal{Q}_{t+2}(x) - \hat{\mathcal{Q}}_{t+2}(x) \right|, \end{aligned} \tag{2.1.137}$$

since $X_{t+1}\left(x_t, \xi_{t+1}\right) \subseteq \mathcal{X}_{t+1}$, for every $x_t \in \mathcal{X}_t$ and $\xi_{t+1} \in E_{t+1} \subseteq \operatorname{supp} \xi_{t+1}$. Summing up, we have shown that

$$\left| \hat{\mathcal{Q}}_{t+1}(x_t) - \mathcal{Q}_{t+1}(x_t) \right| \leq Z_t + \sup_{x \in \mathcal{X}_{t+1}} \left| \mathcal{Q}_{t+2}(x) - \hat{\mathcal{Q}}_{t+2}(x) \right| \tag{2.1.138}$$

holds, for every $x_t \in \mathcal{X}_t$. Taking the supremum in $\mathcal{X}_t$ we obtain that (2.1.135) holds, for every $t = 1, \ldots, T-1$, whenever the event $\mathcal{E}_{N_2,\ldots,N_T}$ happens. It follow that w.p.1

$$\begin{aligned}\sup_{x_1 \in X_1} \left| \mathcal{Q}_2(x_1) - \hat{\mathcal{Q}}_2(x_1) \right| &\leq Z_1 + \cdots + Z_{T-1} + \sup_{x \in \mathbb{R}^{n_T}} \left| \mathcal{Q}_{T+1}(x) - \hat{\mathcal{Q}}_{T+1}(x) \right| \\ &= Z_1 + \cdots + Z_{T-1},\end{aligned}$$

since $\mathcal{Q}_{T+1}(x) = 0 = \hat{\mathcal{Q}}_{T+1}(x)$, for every $x \in \mathbb{R}^{n_T}$.

Note that we can apply Theorem 2.1.5 for every $Z_t$, $t = 1, \ldots, T-1$, since conditions (M1) and (Mt.1)-(Mt.4) are satisfied. Thus, the following bound

$$\mathbb{P}\left[Z_t \geq \frac{\epsilon}{T-1}\right] \leq \exp\left\{-N_{t+1} I_{\chi_t}\left(\tilde{M}_t\right)\right\} + 2\left[\frac{2\rho D_t \tilde{M}_t}{\epsilon/(T-1)}\right]^{n_t} \exp\left\{-\frac{N\epsilon^2}{32\sigma_t^2 (T-1)^2}\right\}, \tag{2.1.139}$$

holds, for every $\epsilon > 0$ and $N_{t+1} \in \mathbb{N}$. Since

$$\begin{aligned}\left[\sup_{x_1 \in X_1} \left| \hat{f}_{N_2,\ldots,N_T}(x_1) - f(x_1) \right| \geq \epsilon\right] \cap \mathcal{E}_{N_2,\ldots,N_T} \subseteq \\ \left(\bigcup_{t=1}^{T-1} \left[Z_t \geq \frac{\epsilon}{T-1}\right]\right) \cap \mathcal{E}_{N_2,\ldots,N_T}\end{aligned}$$

and $\mathbb{P}\left(\mathcal{E}_{N_2,\ldots,N_T}\right) = 1$, it follows that

$$\begin{aligned}\mathbb{P}\left[\sup_{x_1 \in X_1} \left| \hat{f}_{N_2,\ldots,N_T}(x_1) - f(x_1) \right| \geq \epsilon\right] &\leq \mathbb{P}\left(\bigcup_{t=1}^{T-1} \left[Z_t \geq \frac{\epsilon}{T-1}\right]\right) \\ &\leq \sum_{t=1}^{T-1} \mathbb{P}\left[Z_t \geq \frac{\epsilon}{T-1}\right] \\ &\leq \sum_{t=1}^{T-1} \Bigg(\exp\left\{-N_{t+1} I_{\chi_t}\left(\tilde{M}_t\right)\right\} + \\ &\qquad 2\left[\frac{2\rho D_t \tilde{M}_t}{\epsilon/(T-1)}\right]^{n_t} \exp\left\{-\frac{N\epsilon^2}{32\sigma_t^2 (T-1)^2}\right\}\Bigg).\end{aligned}$$

This completes the proof of the proposition. □

Let us make some remarks about Proposition 2.1.16. Note that it was not necessary to assume conditions (Mt.5), for $t = 1, \ldots, T-1$, in order to derive the exponential bound (2.1.127). Moreover, since $f : X_1 \to \mathbb{R}$ is continuous under the hypotheses of Proposition 2.1.16, we have that $\hat{f}_{N_2,\ldots,N_T}(\cdot)$ is bounded in $X_1$, whenever the event

$$\left[\sup_{x_1 \in X_1} \left| \hat{f}_{N_2,\ldots,N_T}(x_1) - f(x_1) \right| < \epsilon\right] \tag{2.1.140}$$

occurs, where $\epsilon > 0$ is arbitrary. So, whenever this event occurs, it automatically follows that $\hat{S}^{\delta}_{N_2,\ldots,N_T} \neq \emptyset$ provided that $\delta > 0$. However, if $\delta = 0$, then it is not clear that $\hat{S}_{N_2,\ldots,N_T} \neq \emptyset$ and some additional regularity conditions such as (Mt.5), $t = 1, \ldots, T-1$, must be assumed in order to guarantee that the SAA problem is solvable.

Akin to the static case, given $\epsilon > 0$, $0 \leq \delta < \epsilon$ and $\theta \in (0,1)$, it is possible to obtain sample complexity estimates for a $T$-stage stochastic programming problem applying Proposition 2.1.16[18]. In [69, 73] sample complexity estimates were derived for arbitrary $\epsilon > 0$ by taking $T = 3$ and $\delta := \epsilon/2 > 0$. In these references, instead of conditions (Mt.4), it was assumed that there exists $M_t \in \mathbb{R}$, $t = 1, \ldots, T-1$, such that

$$|Q_{t+1}(x'_t, \xi_{t+1}) - Q_{t+1}(x_t, \xi_{t+1})| \leq M_t \left\| x'_t - x_t \right\|, \tag{2.1.141}$$

for all $x'_t, x_t \in \mathcal{X}_t$ and w.p.1 $\xi_{t+1}$, $t = 1, 2$. One advantage of assuming these slightly stronger regularity conditions is that, defining $\chi_t(\xi_{T+1}) := M_t$, for $t = 1, \ldots, T-1$, it follows that $I_{\chi_t}(\tilde{M}_t)$ becomes equal to $+\infty$, for every $\tilde{M}_t > M_t$. So, one can simplify the upper bound (2.1.127) getting rid of the terms

$$\exp\left\{-N_{t+1} I_{\chi_t}\left(\tilde{M}_t\right)\right\}, \tag{2.1.142}$$

for $t = 1, \ldots, T-1$, since they vanish in that case. However, we prefer to present the result considering these slightly weaker regularity conditions. Here we also consider the case $\delta = 0$. When we derive in Chapter 5 a lower bound for the sample complexity of a class of $T$-stage stochastic programming problems, it will become clear that these two generalizations are important in order to make a fair comparison between the derived upper and lower bounds for this class of problems.

In Proposition 2.1.17 we obtain the sample complexity estimates for the multistage setting.

**Proposition 2.1.17.** *Take any integer $T \geq 3$. Consider a $T$-stage stochastic programming problem that satisfies conditions (M0), (M1) and (Mt.1)-(Mt.4), for $t = 1, \ldots, T-1$. Denote the sample sizes by $N_2, \ldots, N_T \in \mathbb{N}$ and suppose that the scenario tree is constructed via the* identical conditional sampling *scheme. Let $\tilde{M}_t > M_t := \mathbb{E}\left[\chi_t(\xi_{t+1})\right] \in \mathbb{R}_+$, $t = 1, \ldots, T-1$, be arbitrary real numbers. For $\epsilon > 0$, $\delta \in [0, \epsilon)$ and $\theta \in (0,1)$, define $\tilde{\mathcal{N}}(\epsilon, \delta, \theta) \subseteq \mathbb{N} \times \cdots \times \mathbb{N}$ as:*

$$\left\{ \begin{array}{ll} (N_2, \ldots, N_T) \in \mathbb{N}^{T-1} : & \sum_{t=1}^{T-1} \left( \exp\{-N_{t+1} I_{\chi_t}(\tilde{M}_t)\} + \right. \\ & \left. 2\left[\frac{4\rho D_t \tilde{M}_t}{(\epsilon-\delta)/(T-1)}\right]^{n_t} \exp\left\{-\frac{N_{t+1}(\epsilon-\delta)^2}{128\sigma_t^2 (T-1)^2}\right\} \right) \leq \theta \end{array} \right\} \tag{2.1.143}$$

[18]More precisely, this holds for $0 < \delta < \epsilon$. For $\delta = 0$, the result follows by invoking also Proposition 2.1.15.

*If* $(N_2, \ldots, N_T) \in \tilde{\mathcal{N}}(\epsilon, \delta, \theta)$ *and* $\delta > 0$*, then*

$$\mathbb{P}\left(\left[\hat{S}^{\delta}_{N_2,\ldots,N_T} \subseteq S^{\epsilon}\right] \cap \left[\hat{S}^{\delta}_{N_2,\ldots,N_T} \neq \emptyset\right]\right) \geq 1 - \theta. \tag{2.1.144}$$

*If, additionally, conditions (Mt.5) are satisfied, for* $t = 1, \ldots, T-1$*, then (2.1.144) also holds for* $\delta = 0$*.*

*Proof.* Before proceeding, let us introduce the following (local) notation:

$$E^{\epsilon}_{N_2,\ldots,N_T}(X_1) := \left[\sup_{x_1 \in X_1} \left|\hat{f}_{N_2,\ldots,N_T}(x_1) - f(x_1)\right| < \epsilon\right], \tag{2.1.145}$$

for $\epsilon > 0$. Assume that conditions (M0), (M1) and (Mt.1)-(Mt.4) hold, for $t = 1, \ldots, T-1$. Take any $\epsilon > 0$, $\delta \in [0, \epsilon)$ and $\theta \in (0,1)$. Since $(\epsilon - \delta)/2 > 0$, we have by Proposition 2.1.16 that

$$\mathbb{P}\left(E^{\frac{\epsilon-\delta}{2}}_{N_2,\ldots,N_T}(X_1)\right) \geq 1 - \theta, \tag{2.1.146}$$

whenever $(N_2, \ldots, N_T) \in \tilde{\mathcal{N}}(\epsilon, \delta, \theta)$. Note that

$$E^{\frac{\epsilon-\delta}{2}}_{N_2,\ldots,N_T}(X_1) \subseteq \left[\hat{S}^{\delta}_{N_2,\ldots,N_T} \subseteq S^{\epsilon}\right]. \tag{2.1.147}$$

Indeed, since $X_1$ is a nonempty compact set and $f$ is continuous on $X_1$, it follows that $v^* \in \mathbb{R}$. So, whenever the event $E^{\frac{\epsilon-\delta}{2}}_{N_2,\ldots,N_T}(X_1)$ occurs, it follows from Proposition 2.8.4 that

$$\left|\hat{v}^*_{N_2,\ldots,N_T} - v^*\right| \leq \sup_{x_1 \in X_1} \left|\hat{f}_{N_2,\ldots,N_T}(x_1) - f(x_1)\right| < \frac{\epsilon - \delta}{2}. \tag{2.1.148}$$

So, if $x \in \hat{S}^{\delta}_{N_2,\ldots,N_T}$, then

$$f(x) - \frac{\epsilon - \delta}{2} \leq \hat{f}_{N_2,\ldots,N_T}(x) \leq \hat{v}^*_{N_2,\ldots,N_T} + \delta \leq \left(v^* + \frac{\epsilon - \delta}{2}\right) + \delta, \tag{2.1.149}$$

i.e. $x \in S^{\epsilon}$. This proves that the inclusion (2.1.147) is always satisfied (if $\hat{S}^{\delta}_{N_2,\ldots,N_T} = \emptyset$, then $\hat{S}^{\delta}_{N_2,\ldots,N_T} \subseteq S^{\epsilon}$ is automatically satisfied).

Now, consider that $\delta > 0$. In that case, since $\hat{v}^*_{N_2,\ldots,N_T}$ is finite whenever the event $E^{\frac{\epsilon-\delta}{2}}_{N_2,\ldots,N_T}(X_1)$ happens, it follows that $\hat{S}^{\delta}_{N_2,\ldots,N_T} \neq \emptyset$. We conclude that

$$\begin{aligned}
\mathbb{P}\left(\left[\hat{S}^{\delta}_{N_2,\ldots,N_T} \subseteq S^{\epsilon}\right] \cap \left[\hat{S}^{\delta}_{N_2,\ldots,N_T} \neq \emptyset\right]\right) &\geq \mathbb{P}\left(E^{\frac{\epsilon-\delta}{2}}_{N_2,\ldots,N_T}(X_1)\right) \\
&\geq 1 - \theta,
\end{aligned}$$

whenever $(N_2, \ldots, N_T) \in \tilde{\mathcal{N}}(\epsilon, \delta, \theta)$. Now, suppose additionally that conditions (Mt.5), for $t = 1, \ldots, T-1$, are satisfied. From item (b) of Proposition 2.1.15 we obtain that $\mathbb{P}\left[\hat{S}^{\delta}_{N_2,\ldots,N_T} \neq \emptyset\right] \geq \mathbb{P}\left[\hat{S}_{N_2,\ldots,N_T} \neq \emptyset\right] = 1$, for every $\delta \geq 0$. Therefore, for given $\epsilon > 0$, $\delta \in [0,\epsilon)$ and $\theta \in (0,1)$, we have that

$$\begin{aligned}
\mathbb{P}\left(\left[\hat{S}^{\delta}_{N_2,\ldots,N_T} \subseteq S^{\epsilon}\right] \cap \left[\hat{S}^{\delta}_{N_2,\ldots,N_T} \neq \emptyset\right]\right) &\geq \mathbb{P}\left(E^{\frac{\epsilon-\delta}{2}}_{N_2,\ldots,N_T}(X_1) \cap \left[\hat{S}^{\delta}_{N_2,\ldots,N_T} \neq \emptyset\right]\right) \\
&= \mathbb{P}\left(E^{\frac{\epsilon-\delta}{2}}_{N_2,\ldots,N_T}(X_1)\right) \geq 1-\theta,
\end{aligned}$$

whenever $(N_2,\ldots,N_T) \in \tilde{\mathcal{N}}(\epsilon,\delta,\theta)$. This completes the proof of the proposition. $\square$

In general the number of scenarios of a SAA scenario tree constructed via the *identical conditional sampling* scheme is at most equal to

$$\prod_{t=2}^{T} N_t, \tag{2.1.150}$$

where $N_2,\ldots,N_T \in \mathbb{N}$ are the number of samples taken from each random vector $\xi_t$, for $t = 2,\ldots,T$. We have already pointed that if each random vector $\xi_t$, $t = 2,\ldots,T$, has a continuous marginal distribution[19], then the number of scenarios is exactly equal to (2.1.150) w.p.1. Thus, in the multistage setting, we study how quantity (2.1.150) grows with respect to: (a) the sample complexity parameters $\epsilon > 0$, $\delta \in [0,\epsilon)$, $\theta \in (0,1)$, (b) the problem data such as $M_t$, $\sigma_t^2$ and $D_t$, for $t = 1,\ldots,T-1$, and (c) the number of stages $T \geq 3$.

Consider the following quantity

$$\tilde{N}(\epsilon,\delta,\theta) := \inf\left\{\prod_{t=2}^{T} N_t : (N_2,\ldots,N_T) \in \tilde{\mathcal{N}}(\epsilon,\delta,\theta)\right\}, \tag{2.1.151}$$

where $\epsilon > 0$, $\delta \in [0,\epsilon)$, and $\theta \in (0,1)$. Since $\tilde{\mathcal{N}}(\epsilon,\delta,\theta) \neq \emptyset$[20], the infimum in (2.1.151) is achieved for some $(N_2^*,\ldots,N_T^*) \in \tilde{\mathcal{N}}(\epsilon,\delta,\theta)$. In fact, the set in (2.1.151) is discrete, since it is a subset of $\mathbb{N}$. The quantity $\tilde{N}(\epsilon,\delta,\theta)$ is an upper bound for the sample complexity of $T$-stage stochastic programming problems satisfying the stated regularity conditions. Maybe it is not easy to obtain a closed formula expression for $\tilde{N}(\epsilon,\delta,\theta)$; however it is not difficult to derive lower and upper estimates for this quantity (see Lemma 2.1.18). In the sequel, we use these estimates in order to study the behavior of $\tilde{N}(\epsilon,\delta,\theta)$ with respect to the sample complexity parameters, the problem data and the number of stages.

[19]Let us recall that we are assuming here the stagewise independent hypothesis.

[20]In fact, one just needs to take each $N_{t+1}$ sufficiently large, since $I_{\chi_t}(\tilde{M}_t) > 0$ and $\frac{(\epsilon-\delta)^2}{128\sigma_t^2(T-1)^2} > 0$, for $t = 1,\ldots,T-1$.

**Lemma 2.1.18.** *Take any integer $T \geq 3$. Given $\theta \in (0,1)$, consider the following subset of $\mathbb{N}^{T-1}$:*

$$\mathfrak{N} := \left\{ \ (N_2, \dots, N_T) \in \mathbb{N}^{T-1} : \ \sum_{t=1}^{T-1} \left[\exp\left\{-\beta_{1,t} N_{t+1}\right\} + C_t \exp\left\{-\beta_{2,t} N_{t+1}\right\}\right] \leq \theta \ \right\}, \tag{2.1.152}$$

*where $\beta_{1,t}$, $\beta_{2,t}$ and $C_t$ are positive real numbers, for $t = 1, \dots, T-1$. Define the following quantity*

$$N := \inf \left\{ \prod_{t=2}^{T} N_t : (N_2, \dots, N_T) \in \mathfrak{N} \right\}. \tag{2.1.153}$$

*We have that $\underline{\mathfrak{N}} \subseteq \mathfrak{N} \subseteq \overline{\mathfrak{N}}$, where*

$$\begin{aligned} \underline{\mathfrak{N}} \ &:= \ \left\{ (N_2, \dots, N_T) : \begin{array}{l} \exp\{-\beta_{1,t} N_{t+1}\} \leq \frac{\theta}{2(T-1)} \ \textit{and} \\ C_t \exp\{-\beta_{2,t} N_{t+1}\} \leq \frac{\theta}{2(T-1)}, \ \textit{for } 1 \leq t \leq T-1 \end{array} \right\}, \\ \overline{\mathfrak{N}} \ &:= \ \left\{ (N_2, \dots, N_T) : \begin{array}{l} \exp\{-\beta_{1,t} N_{t+1}\} \leq \theta \ \textit{and} \ C_t \exp\{-\beta_{2,t} N_{t+1}\} \leq \theta, \\ \textit{for } t = 1, \dots, T-1 \end{array} \right\}. \end{aligned}$$

*In particular,*

$$\inf \left\{ \prod_{t=2}^{T} N_t : (N_2, \dots, N_T) \in \overline{\mathfrak{N}} \right\} \leq N \leq \inf \left\{ \prod_{t=2}^{T} N_t : (N_2, \dots, N_T) \in \underline{\mathfrak{N}} \right\}. \tag{2.1.154}$$

*Proof.* Take any $(N_2, \dots, N_T) \in \underline{\mathfrak{N}}$. It follows that

$$\sum_{t=1}^{T-1} \left[\exp\left\{-\beta_{1,t} N_{t+1}\right\} + C_t \exp\left\{-\beta_{2,t} N_{t+1}\right\}\right] \leq \sum_{t=1}^{T-1} \left[\frac{\theta}{2(T-1)} + \frac{\theta}{2(T-1)}\right] = \theta, \tag{2.1.155}$$

i.e. $(N_2, \dots, N_T) \in \mathfrak{N}$.

Now, take any $(N_2, \dots, N_T) \in \mathfrak{N}$. Since each term $\exp\left\{-\beta_{1,t} N_{t+1}\right\}$ and $C_t \exp\{-\beta_{2,t} N_{t+1}\}$ is nonnegative, for $t = 1, \dots, T-1$, and their sum is less than or equal to $\theta$, it follows that each term is also less than or equal to $\theta$, i.e. $(N_2, \dots, N_T) \in \overline{\mathfrak{N}}$. This completes the proof of the inclusions $\underline{\mathfrak{N}} \subseteq \mathfrak{N} \subseteq \overline{\mathfrak{N}}$. Equation (2.1.154) follows immediately from these inclusions. $\square$

Note that it is elementary to obtain closed form formulae for $\inf \left\{ \prod_{t=2}^{T} N_t : (N_2, \dots, N_T) \in \overline{\mathfrak{N}} \right\}$ and $\inf \left\{ \prod_{t=2}^{T} N_t : (N_2, \dots, N_T) \in \underline{\mathfrak{N}} \right\}$. In fact, $(N_2, \dots, N_T) \in \overline{\mathfrak{N}}$ if and only if

$$N_{t+1} \geq \left\lceil \frac{1}{\beta_{1,t}} \log\left(\frac{1}{\theta}\right) \right\rceil \vee \left\lceil \frac{1}{\beta_{2,t}} \log\left(\frac{C_t}{\theta}\right) \right\rceil, \tag{2.1.156}$$

for $t = 1, \ldots, T-1$. Therefore,

$$\inf\left\{\prod_{t=2}^{T} N_t : (N_2, \ldots, N_T) \in \overline{\mathfrak{N}}\right\} = \prod_{t=1}^{T-1}\left(\left\lceil \frac{1}{\beta_{1,t}} \log\left(\frac{1}{\theta}\right)\right\rceil \vee \left\lceil \frac{1}{\beta_{2,t}} \log\left(\frac{C_t}{\theta}\right)\right\rceil\right). \tag{2.1.157}$$

Similarly, $(N_2, \ldots, N_T) \in \underline{\mathfrak{N}}$ if and only if

$$N_{t+1} \geq \left\lceil \frac{1}{\beta_{1,t}} \log\left(\frac{2(T-1)}{\theta}\right)\right\rceil \vee \left\lceil \frac{1}{\beta_{2,t}} \log\left(\frac{2(T-1)C_t}{\theta}\right)\right\rceil, \tag{2.1.158}$$

for $t = 1, \ldots, T-1$. Therefore,

$$\begin{aligned}\inf\left\{\prod_{t=2}^{T} N_t : (N_2, \ldots, N_T) \in \underline{\mathfrak{N}}\right\} = \prod_{t=1}^{T-1}&\left(\left\lceil \tfrac{1}{\beta_{1,t}} \log\left(\tfrac{2(T-1)}{\theta}\right)\right\rceil \vee\right.\\ &\left.\left\lceil \tfrac{1}{\beta_{2,t}} \log\left(\tfrac{2(T-1)C_t}{\theta}\right)\right\rceil\right).\end{aligned} \tag{2.1.159}$$

Now we discuss the implications of Lemma 2.1.18 on the sample complexity estimate for $T$-stage stochastic programming problems. Consider a $T$-stage problem satisfying conditions (M0), (M1), (Mt.1)-(Mt.5), for $t = 1, \ldots, T-1$. Let $\tilde{M}_t > M_t$ be given real numbers and consider the sample complexity parameters $\epsilon > 0$, $\delta \in [0, \epsilon)$ and $\theta \in (0, 1)$. For such values, we have that

$$\beta_{1,t} = I_{\chi_t}\left(\tilde{M}_t\right), \tag{2.1.160}$$

$$\beta_{2,t} = \frac{(\epsilon-\delta)^2}{128\sigma_t^2(T-1)^2}, \text{ and} \tag{2.1.161}$$

$$C_t = 2\left[\frac{4\rho D_t \tilde{M}_t (T-1)}{\epsilon - \delta}\right]^{n_t}, \tag{2.1.162}$$

for $t = 1, \ldots, T-1$.

Now, note that by taking $T = 2$ and using equation (2.1.156) we obtain that $\tilde{N}(\epsilon, \delta, \theta)$ is greater than or equal to

$$\left[\frac{1}{I_{\chi_1}(\tilde{M}_1)} \log\left(\frac{1}{\theta}\right)\right] \vee \left(\frac{128\sigma_1^2}{(\epsilon-\delta)^2}\left[n_1 \log\left(\frac{4\rho D_1 \tilde{M}_1}{\epsilon-\delta}\right) + \log\left(\frac{2}{\theta}\right)\right]\right). \tag{2.1.163}$$

This recovers the sample complexity estimate obtained for 2-stage stochastic programming problems (see estimate (2.1.46)). As discussed previously, for sufficiently small values of $\epsilon - \delta > 0$ the second term of the maximum (2.1.163) is greater than its first term. In that case, the sample complexity estimate for 2-stage problems is just

$$N_2 \geq \left(\frac{128\sigma_1^2}{(\epsilon-\delta)^2}\left[n_1 \log\left(\frac{4\rho D_1 \tilde{M}_1}{\epsilon-\delta}\right) + \log\left(\frac{2}{\theta}\right)\right]\right). \tag{2.1.164}$$

In order to fix some ideas, we assume that

$$\sigma_1 = \sigma_2 = \cdots = \sigma_{T-1}, \tag{2.1.165}$$
$$D_1\tilde{M}_1 = D_2\tilde{M}_2 = \cdots = D_{T-1}\tilde{M}_{T-1}, \tag{2.1.166}$$
$$n_1 = n_2 = \cdots = n_{T-1}. \tag{2.1.167}$$

Using again equation (2.1.156) we obtain that

$$\begin{aligned}\tilde{N}(\epsilon,\delta,\theta) &\geq \prod_{t=1}^{T-1}\left(\tfrac{128\sigma_t^2(T-1)^2)}{(\epsilon-\delta)^2}\left[n_t\log\left(\tfrac{4\rho D_t\tilde{M}_t(T-1)}{\epsilon-\delta}\right)+\log\left(\tfrac{2}{\theta}\right)\right]\right)\\ &\geq (T-1)^{2(T-1)}\left(\tfrac{128\sigma_1^2}{(\epsilon-\delta)^2}\left[n_1\log\left(\tfrac{4\rho D_1\tilde{M}_1}{\epsilon-\delta}\right)+\log\left(\tfrac{2}{\theta}\right)\right]\right)^{T-1}\end{aligned} \tag{2.1.168}$$

Therefore, the sample complexity estimate for $T$-stage problems is like the estimate obtained for 2-stage problems to the power of $T-1$ multiplied by the factor

$$(T-1)^{2(T-1)}.$$

Treating the number of stages $T$ as a varying parameter, it follows that (2.1.168) has an order of growth with respect to the $T$ that is much greater than simply the estimate obtained for static problems to the power of $T-1$. Indeed, the factor $(T-1)^{2(T-1)}$ grows even faster than the factorial function $T!$ with respect to $T$. Recall that the factorial function grows much faster than the exponential function $c^T$, for $c>1$ constant. It is worth mentioning that in Chapter 5 (see also [53]) we show that this is an unavoidable phenomenon for some $T$-stage stochastic programming problems. In fact, we have shown that even some problems satisfying "nice" regularity conditions such as the ones considered here can present this kind of behavior with respect to $T$.

Now, let us consider how the sample complexity parameters $\epsilon$, $\delta$ and $\theta$ affects estimate $\tilde{N}(\epsilon,\delta,\theta)$. $\tilde{N}(\epsilon,\delta,\theta)$ depends on $\epsilon$ and $\delta$ only through the difference $\epsilon-\delta>0$. So, without loss of generality, take $\delta=0$. Of course, the sample complexity estimate grows whenever we ask for obtaining a more accurate solution of the true problem, i.e. for smaller values of $\epsilon>0$. Then, *ceteris paribus*[21],

$$\tilde{N}(\epsilon) = \mathrm{O}\left(\frac{1}{\epsilon^{2(T-1)}}\left[\log\left(\frac{1}{\epsilon}\right)\right]^{T-1}\right) \tag{2.1.169}$$

when $\epsilon>0$ approaches 0. Furthermore, note that the sample complexity grows whenever we ask for obtaining an approximate solution with higher degree of certainty, i.e. for $\theta>0$ small. The dependence of $\tilde{N}(\epsilon,\delta,\theta)$ with respect to $\theta$ is of order

$$\tilde{N}(\theta) = \mathrm{O}\left(\log\left(\frac{1}{\theta}\right)^{T-1}\right), \tag{2.1.170}$$

[21] That is, considering the remaining parameters fixed.

*ceteris paribus.*

It is worth noting that assumptions (2.1.165), (2.1.166), and (2.1.167) are not essential for obtaining this order of growth. In fact, we can apply the same reasoning by considering in (2.1.168) the following quantities

$$\sigma := \min_{1\leq t\leq T-1} \sigma_t, \tag{2.1.171}$$
$$D := \min_{1\leq t\leq T-1} D_t, \tag{2.1.172}$$
$$\tilde{M} := \min_{1\leq t\leq T-1} \tilde{M}_t, \tag{2.1.173}$$
$$n := \min_{1\leq t\leq T-1} n_t, \tag{2.1.174}$$

respectively, instead of considering $\sigma_1$, $D_1$, $\tilde{M}_1$, and $n_1$.

In Chapter 4 we extend the analysis done here for multistage stochastic programming problems with nested OCE risk measures.

## 2.2 Scenario Trees

In this section we make a detailed exposition of scenario trees that are objects commonly used by the stochastic programming community for representing finite state space stochastic processes $\xi = (\xi_1, \ldots, \xi_T)$, where $T \geq 2$.

Let us begin by defining a directed rooted tree.

**Definition 2.2.1.** *(directed rooted trees) A directed rooted tree is a tree $(\mathcal{N}, \mathcal{A})$ in which a node $\iota_1$ was selected as a root node, $\mathcal{N}$ is the set of nodes of the tree, and $\mathcal{A} \subseteq \mathcal{N} \times \mathcal{N}$ is the set of arcs of the tree. Here we always assume that every arc of the tree points away from the root $\iota_1$.*

We always assume that $\mathcal{N} \neq \emptyset$ is finite. Now, let us present some standard terminology concerning trees. Take any directed rooted tree $(\mathcal{N}, \mathcal{A})$ and let $\iota_1 \in \mathcal{N}$ be its root node. $\iota_1$ is the unique node of the tree at level $t = 1$, i.e. $\mathcal{N}_1 = \{\iota_1\}$, where $\mathcal{N}_t$ is the set of nodes of the tree at level $t \in \mathbb{N}$. For each node $\iota$ we consider the set $C_\iota \subseteq \mathcal{N}$ of children of $\iota$

$$C_\iota := \{\iota' \in \mathcal{N} : (\iota, \iota') \in \mathcal{A}\}.$$

If $C_\iota = \emptyset$, then we say that $\iota$ is a leaf node of the tree or a terminal node. For $t \geq 2$, the set of nodes $\mathcal{N}_t$ at level $t$ of the tree is given by

$$\mathcal{N}_t := \{\iota_t \in \mathcal{N} : \text{for } s = 1, \ldots, t-1 \text{ there exists } \iota_{s+1} \in C_{\iota_s}\}. \tag{2.2.1}$$

This means that the (unique) path $(\iota_1, \iota_2, \ldots, \iota_t)$ connecting the root node $\iota_1$ with node $\iota_t$ has length $t$. Since $\mathcal{N}$ is finite, $\mathcal{N}_t = \emptyset$, for $t > \operatorname{card}\mathcal{N}$. The depth of the tree is given by

$$T := \max\{t \in \mathbb{N} : \mathcal{N}_t \neq \emptyset\}. \tag{2.2.2}$$

Note that

$$\mathcal{N} = \bigcup_{t=1}^{T} \mathcal{N}_t, \tag{2.2.3}$$

and $\mathcal{N}_t \cap \mathcal{N}_s = \emptyset$, for $s \neq t$. Except when otherwise stated, we assume that the path connecting the root node to any leaf node has length $T \geq 2$. Equivalently, we have that $C_\iota \neq \emptyset$, for all $\iota \in \mathcal{N}_t$, and $1 \leq t < T$. The following equalities are also satisfied

$$\mathcal{N}_1 = \{\iota_1\},\ \mathcal{N}_2 = C_{\iota_1},\ \mathcal{N}_3 = \bigcup_{\iota \in \mathcal{N}_2} C_\iota,\ \ldots,\ \mathcal{N}_T = \bigcup_{\iota \in \mathcal{N}_{T-1}} C_\iota. \tag{2.2.4}$$

Every node $\iota \neq \iota_1$ has a unique parent node $a(\iota)$ that is characterized as the unique node that satisfies $(a(\iota), \iota) \in \mathcal{A}$. Note also that $\iota \in C_{a(\iota)}$.

In order to introduce scenario trees suppose that each node $\iota \in \mathcal{N}_t$ is related to a vector $\xi_t^\iota \in \mathbb{R}^{d_t}$. We also assume that for each $\iota, \iota' \in \mathcal{N}_t$

$$\xi_s^{a^{t-s}(\iota)} \neq \xi_s^{a^{t-s}(\iota')}, \tag{2.2.5}$$

for some $2 \leq s \leq t \leq T$, if $\iota \neq \iota'$. For clarifying the notation, we define $a^0 := \mathrm{Id}$ and $a^s := a \circ a^{s-1}$, for $s = 1, \ldots, T$. This condition means that every node $\iota_t$ at level $t$ can be distinguished from a different node $\iota_t'$ at level $t$ by looking at the sequences of values associated with each node of the path that connect the root node to each one of these nodes.

Let us also consider a family of positive[22] numbers $\rho := \{\rho_a : a \in \mathcal{A}\}$ defined on $\mathcal{A}$ that satisfies

$$\sum_{\iota' \in C_\iota} \rho_{(\iota, \iota')} = 1 \tag{2.2.6}$$

for every $\iota \in \mathcal{N}_t$, and $t = 1, \ldots, T-1$. The set of all paths from the root node to the leaves nodes is

$$\mathcal{S} := \{(\iota_1, \ldots, \iota_T) \in \mathcal{N}_1 \times \cdots \times \mathcal{N}_T : (\iota_t, \iota_{t+1}) \in \mathcal{A},\ \text{for } t = 1, \ldots, T-1\}. \tag{2.2.7}$$

Using $\rho := \{\rho_a : a \in \mathcal{A}\}$ we can associate a probability value for each complete path[23] or scenario of the tree

$$\rho_{(\iota_1, \ldots, \iota_T)} := \prod_{t=1}^{T-1} \rho_{(\iota_t, \iota_{t+1})}. \tag{2.2.8}$$

It is instructive to show that $\{\rho_{(\iota_1, \ldots, \iota_T)} : (\iota_1, \ldots, \iota_T) \in \mathcal{S}\}$ really defines a probability on $\mathcal{S}$[24] Since $\mathcal{S}$ is finite, we just need to verify that

$$\rho_{(\iota_1, \ldots, \iota_T)} \geq 0,\ \forall (\iota_1, \ldots, \iota_T) \in \mathcal{S},\ \text{and} \tag{2.2.9}$$

$$\sum_{(\iota_1, \ldots, \iota_T) \in \mathcal{S}} \rho_{(\iota_1, \ldots, \iota_T)} = 1. \tag{2.2.10}$$

[22]We assume that $\rho_a > 0$, for every $a \in \mathcal{A}$.

[23]In the sense that the path begins at the root node of the tree and travels until one of its leaf nodes.

[24]Or more precisely, on $\mathcal{P}(\mathcal{S})$.

The first condition follows immediately from (2.2.8), since $\rho_{\iota_t,\iota_{t+1}} > 0$, for every $t = 1, \ldots, T-1$, whenever $(\iota_1, \ldots, \iota_T) \in \mathcal{S}$. We can also write $\mathcal{S}$ in the following way

$$\mathcal{S} = \left\{(\iota_1, \ldots, \iota_T) \in \mathcal{N}_1 \times \cdots \times \mathcal{N}_T : \iota_2 \in C_{\iota_1}, \iota_3 \in C_{\iota_2}, \ldots, \iota_T \in C_{\iota_{T-1}}\right\}. \quad (2.2.11)$$

Therefore, the sum (2.2.10) can also be written as the iterated sum

$$\begin{aligned}
\sum_{(\iota_1,\ldots,\iota_T)\in\mathcal{S}} \rho_{(\iota_1,\ldots,\iota_T)} &= \sum_{\iota_2\in C_{\iota_1}} \sum_{\iota_3\in C_{\iota_2}} \cdots \sum_{\iota_T\in C_{\iota_{T-1}}} \rho_{(\iota_1,\ldots,\iota_T)} && (2.2.12)\\
&= \sum_{\iota_2\in C_{\iota_1}} \sum_{\iota_3\in C_{\iota_2}} \cdots \sum_{\iota_T\in C_{\iota_{T-1}}} \rho_{(\iota_1,\iota_2)} \cdots \rho_{(\iota_{T-1},\iota_T)} && (2.2.13)\\
&= \sum_{\iota_2\in C_{\iota_1}} \rho_{(\iota_1,\iota_2)} \cdots \sum_{\iota_{T-1}\in C_{\iota_{T-2}}} \rho_{(\iota_{T-2},\iota_{T-1})} \sum_{\iota_T\in C_{\iota_{T-1}}} \rho_{(\iota_{T-1},\iota_T)} && (2.2.14)\\
&= \sum_{\iota_2\in C_{\iota_1}} \rho_{(\iota_1,\iota_2)} \cdots \sum_{\iota_{T-1}\in C_{\iota_{T-2}}} \rho_{(\iota_{T-2},\iota_{T-1})} && (2.2.15)\\
&\;\;\vdots && (2.2.16)\\
&= \sum_{\iota_2\in C_{\iota_1}} \rho_{(\iota_1,\iota_2)} = 1, && (2.2.17)
\end{aligned}$$

using that $\sum_{\iota_t\in C_{\iota_{t-1}}} \rho_{(\iota_{t-1},\iota_t)} = 1$, for every $\iota_{t-1} \in \mathcal{N}_{t-1}$ and $t = T, \ldots, 2$.

The value $\rho_{(\iota_t,\iota_{t+1})}$ represents the conditional probability of going from node $\iota_t \in \mathcal{N}_t$ to node $\iota_{t+1} \in C_\iota \subseteq \mathcal{N}_{t+1}$ given that we are currently at node $\iota_t$. We are ready to present the definition of a scenario tree that we consider in this thesis.

**Definition 2.2.2.** *(scenario trees) A scenario tree $\tau := (\mathcal{N}, \mathcal{A}, \xi, \rho)$ is a tuple satisfying the following conditions:*

*(i) $(\mathcal{N}, \mathcal{A})$ is a directed rooted tree with root node $\iota_1$ and such that every arc of the tree points away from the root node.*

*(ii) the family of vectors $\xi = \{\xi_t^\iota \in \mathbb{R}^{n_t} : \iota \in \mathcal{N}_t, t = 1, \ldots, T\}$ is such that condition (2.2.5) is satisfied.*

*(iii) the family of positive numbers $\rho := \{\rho_a : a \in \mathcal{A}\}$ is such that condition (2.2.6) is satisfied.*

Given a scenario tree $(\mathcal{N}, \mathcal{A}, \xi, \rho)$, we consider, with a slight abuse of notation, its associated stochastic process defined on the probability space $(\mathcal{S}, \mathcal{P}(\mathcal{S}), \{\rho_s : s \in \mathcal{S}\})$ of scenarios of the tree:

$$\begin{array}{rccl}
\xi : & \mathcal{S} & \to & \mathbb{R}^d = \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_T} \\
& (\iota_1, \ldots, \iota_T) & \mapsto & (\xi_1^{\iota_1}, \ldots, \xi_T^{\iota_T})
\end{array} \quad (2.2.18)$$

The random vector $\xi_t$ of stage $t$ is just the projection of $\xi$ on $\mathbb{R}^{d_t}$, for $t = 1, \ldots, T$.

When we describe sampling schemes for approximating the true random data by a finite state space stochastic process in Section 2.1.2, we detail a procedure that generates a scenario tree. Then, the SAA stochastic process is just the stochastic process associated to the generated scenario tree.

## 2.3 Quantiles

**Definition 2.3.1.** *Let $Z$ be a random variable and $\alpha \in [0,1]$. We say that $z \in \mathbb{R}$ is an $\alpha$-quantile of $Z$ if the following conditions hold:*

*(a)* $\mathbb{P}[Z \leq z] \geq \alpha$,

*(b)* $\mathbb{P}[Z \geq z] \geq 1 - \alpha$.

The following result is well-known, but we present a proof for the sake of self-containedness.

**Proposition 2.3.2.** *Let $Z$ be a random variable and $\alpha \in (0,1)$ be arbitrary. Define the following quantities:*

$$\begin{aligned} q_\alpha^-(Z) &:= \inf\{z \in \mathbb{R} : \mathbb{P}[Z \leq z] \geq \alpha\}, \text{ and} && (2.3.1)\\ q_\alpha^+(Z) &:= \sup\{z \in \mathbb{R} : \mathbb{P}[Z \geq z] \geq 1-\alpha\}. && (2.3.2)\end{aligned}$$

*We have that $q_\alpha^-(Z) \leq q_\alpha^+(Z)$ are finite numbers and that the set of $\alpha$-quantiles of $Z$ is given by the (nonempty compact) interval $[q_\alpha^-(Z), q_\alpha^+(Z)]$.*

*Proof.* Let $\alpha \in (0,1)$ be given and denote by:

$$\begin{aligned} I_\alpha^- &:= \{z \in \mathbb{R} : \mathbb{P}[Z \leq z] \geq \alpha\}, && (2.3.3)\\ I_\alpha^+ &:= \{z \in \mathbb{R} : \mathbb{P}[Z \geq z] \geq 1-\alpha\}. && (2.3.4)\end{aligned}$$

We will show that $I_\alpha^-$ and $I_\alpha^+$ are closed unbounded intervals that are, respectively, bounded from below and above. First of all, observe that if $z \in I_\alpha^-$ and $w > z$, then $\mathbb{P}[Z \leq w] \geq \mathbb{P}[Z \leq z] \geq \alpha$, i.e. $w \in I_\alpha^-$. Moreover, we have that:

$$\begin{aligned} \Omega &= \bigcup_{k\in\mathbb{N}} [Z \leq k], \text{ and} && (2.3.5)\\ \emptyset &= \bigcap_{k\in\mathbb{N}} [Z \leq -k]. && (2.3.6)\end{aligned}$$

It follows that $\lim_{k\to+\infty} \mathbb{P}[Z \leq k] = 1 > \alpha > 0 = \lim_{k\to+\infty} \mathbb{P}[Z \leq -k]$. Therefore, $I_\alpha^- \neq \emptyset$ is an unbounded interval that is bounded from below. Let us show that $q_\alpha^-(Z) \in I_\alpha^-$. We have that $q_\alpha^-(Z) + 1/k \in I_\alpha^-$, for all $k \in \mathbb{N}$. Moreover,

$$[Z \leq q_\alpha^-(Z)] = \bigcap_{k\in\mathbb{N}} [Z \leq q_\alpha^-(Z) + 1/k] \quad (2.3.7)$$

which implies that: $\mathbb{P}[Z \leq q_\alpha^-(Z)] = \lim_{k\to+\infty} \mathbb{P}[Z \leq q_\alpha^-(Z) + 1/k] \geq \alpha$. This shows that $I_\alpha^- = [q_\alpha^-(Z), +\infty)$. In order to show that $I_\alpha^+ = (-\infty, q_\alpha^+(Z)]$ we follow similar steps. Omitting some details, let us just observe that:

$$\Omega = \bigcup_{k\in\mathbb{N}} [Z \geq -k]; \tag{2.3.8}$$

$$\emptyset = \bigcap_{k\in\mathbb{N}} [Z \geq k]; \tag{2.3.9}$$

if $z \in I_\alpha^+$ and $w < z$, then $w \in I_\alpha^+$; $Z \geq q_\alpha^+(Z) = \bigcap_{k\in\mathbb{N}} [Z \geq q_\alpha^+(Z) - 1/k]$, and the argument follows similarly.

Now, let us show that $q_\alpha^-(Z) \leq q_\alpha^+(Z)$. For $z < q_\alpha^-(Z)$, we have that $\mathbb{P}[Z \leq z] < \alpha$. Therefore, $\mathbb{P}[Z \geq z] = 1 - \mathbb{P}[Z < z] \geq 1 - \mathbb{P}[Z \leq z] > 1 - \alpha$. We conclude that:

$$(-\infty, q_\alpha^-(Z)) \subseteq I_\alpha^+(Z),$$

so: $q_\alpha^-(Z) \leq q_\alpha^+(Z)$. This also shows that $[q_\alpha^-(Z), q_\alpha^+(Z)] \neq \emptyset$. Finally, observe that $z$ is an $\alpha$-quantile of $Z$ if and only if $z \in I_\alpha^- \cap I_\alpha^+ = [q_\alpha^-(Z), q_\alpha^+(Z)]$. The proposition is proved. □

Proposition 2.3.2 states that $q_\alpha^-(Z)$ is the *minimum* or the *leftmost* $\alpha$-quantile of $Z$, $q_\alpha^+(Z)$ is the *maximum* or the *rightmost* $\alpha$-quantile of $Z$, and that these quantities are finite when $\alpha \in (0, 1)$. When $\alpha = 0$ or $\alpha = 1$ the set of $\alpha$-quantiles of a random variable can be the empty set or an unbounded interval. The following lemma will be useful for showing some interesting properties of the leftmost and rightmost $\alpha$-quantiles functions of a given random variable.

**Lemma 2.3.3.** *Take a random variable $Z$ and $\alpha \in (0, 1)$ . The leftmost and rightmost $\alpha$-quantiles of $Z$ admit the following alternatives characterization:*

$$q_\alpha^-(Z) = \sup\{z \in \mathbb{R} : \mathbb{P}[Z < z] < \alpha\}, \tag{2.3.10}$$

$$q_\alpha^+(Z) = \inf\{z \in \mathbb{R} : \mathbb{P}[Z > z] < 1 - \alpha\}, \tag{2.3.11}$$

*respectively.*

*Proof.* We will show only that equation (2.3.10) holds, since we will only need it in the next proposition. Let us denote by $\tilde{z} := \sup\{z \in \mathbb{R} : \mathbb{P}[Z < z] < \alpha\}$. For an arbitrary $z > \tilde{z}$, we have that $z \notin \{z \in \mathbb{R} : \mathbb{P}[Z < z] < \alpha\}$, that is: $\alpha \leq \mathbb{P}[Z < z] \leq \mathbb{P}[Z \leq z]$. It follows that:

$$(\tilde{z}, +\infty) \subseteq \{z \in \mathbb{R} : \mathbb{P}[Z \leq z] \geq \alpha\}. \tag{2.3.12}$$

Taking the infimum of these sets, we obtain that $q_\alpha^-(Z) \leq \tilde{z}$. Now, observe that:

$$[Z \leq \tilde{z}] = \bigcap_{k\in\mathbb{N}} [Z \leq \tilde{z} + 1/k]. \tag{2.3.13}$$

It follows that:

$$\mathbb{P}[Z \le \tilde{z}] = \lim_{k\to+\infty} \mathbb{P}[Z \le \tilde{z} + 1/k] \ge \alpha. \tag{2.3.14}$$

We conclude that $q_\alpha^-(Z) \le \tilde{z}$ and equation (2.3.10) is proved. The proof of equation (2.3.11) is similar. □

**Proposition 2.3.4.** *Let $Z$ be a given random variable. We have that the rightmost quantile function:*

$$\alpha \in (0,1) \mapsto q_\alpha^+(Z) \in \mathbb{R}, \tag{2.3.15}$$

*and the leftmost quantile function:*

$$\alpha \in (0,1) \mapsto q_\alpha^-(Z) \in \mathbb{R} \tag{2.3.16}$$

*are monotonically non-decreasing and satisfy the following conditions:*

*(i)* $q_\alpha^+(Z) \le q_\beta^-(Z)$, *for all* $0 < \alpha < \beta < 1$;

*(ii) for all* $\alpha \in (0,1)$,

$$q_\alpha^+(Z) = \inf_{\beta>\alpha} q_\beta^+(Z) = \inf_{\beta>\alpha} q_\beta^-(Z); \tag{2.3.17}$$

*(iii) for all* $\beta \in (0,1)$,

$$q_\beta^-(Z) = \sup_{\alpha<\beta} q_\alpha^-(Z) = \inf_{\alpha<\beta} q_\alpha^+(Z). \tag{2.3.18}$$

*Proof.* We begin by showing item (i). Let $0 < \alpha < \beta < 1$ be given. Observe that:

$$\{z \in \mathbb{R} : \mathbb{P}[Z \ge z] \ge 1-\beta\} \subseteq \{z \in \mathbb{R} : \mathbb{P}[Z < z] < \alpha\}. \tag{2.3.19}$$

Indeed, if $\mathbb{P}[Z \ge z] \ge 1-\beta$, then $\mathbb{P}[Z < z] = 1 - \mathbb{P}[Z \ge z] \le 1-(1-\beta) = \beta < \alpha$. Taking the supremum of these sets, it follows from Lemma 2.3.3 that: $q_\beta^+(Z) \le q_\alpha^-(Z)$. We also have that:

$$q_\alpha^-(Z) \le q_\alpha^+(Z) \le q_\beta^-(Z) \le q_\beta^+(Z), \tag{2.3.20}$$

that is, both functions are monotonically non-decreasing.

Now we show item (ii). From item (i), it follows that:

$$q_\alpha^+(Z) \le \inf_{\beta>\alpha} q_\beta^-(Z) \le \inf_{\beta>\alpha} q_\beta^+(Z). \tag{2.3.21}$$

We need only to show that $\inf_{\beta>\alpha} q_\beta^+(Z) \le q_\alpha^+(Z)$. Let $\epsilon > 0$ be given. We have that:

$$\mathbb{P}[Z \ge q_\alpha^+(Z) + \epsilon] < 1-\alpha. \tag{2.3.22}$$

Taking $\beta > \alpha$ sufficiently close to $\alpha$, we obtain that:

$$\mathbb{P}[Z \ge q_\alpha^+(Z) + \epsilon] < 1-\beta \le \mathbb{P}[Z \ge q_\beta^+(Z)], \tag{2.3.23}$$

i.e., $q_\beta^+(Z) < q_\alpha^+(Z) + \epsilon$. It follows that $\inf_{\beta>\alpha} q_\beta^+(Z) \le q_\alpha^+(Z) + \epsilon$, for all $\epsilon > 0$, which proves item (ii). The proof of item (iii) is similar. □

**Proposition 2.3.5.** *Let $Z$ be a random variable and $\alpha \in (0,1)$ be arbitrary. Then,*

$$q_{\alpha}^{-}(Z) = -q_{1-\alpha}^{+}(-Z). \tag{2.3.24}$$

*Proof.* Observe that:

$$\alpha = 1-(1-\alpha) \leq \mathbb{P}\left[-Z \geq q_{1-\alpha}^{+}(-Z)\right] = \mathbb{P}\left[Z \leq -q_{1-\alpha}^{+}(-Z)\right], \tag{2.3.25}$$

so, $q_{\alpha}^{-}(Z) \leq -q_{1-\alpha}^{+}(-Z)$. Moreover,

$$1-(1-\alpha) = \alpha \leq \mathbb{P}\left[Z \geq q_{\alpha}^{-}(Z)\right] = \mathbb{P}\left[-Z \geq -q_{\alpha}^{-}(Z)\right], \tag{2.3.26}$$

so, $q_{1-\alpha}^{+}(-Z) \geq -q_{\alpha}(Z)$, that is, $-q_{1-\alpha}^{+}(-Z) \leq q_{\alpha}^{-}(Z)$ and the result is proved. □

## 2.4 Sub-Gaussian and $\psi_2$-random variables

In this section, we recall the definitions of sub-Gaussian and $\psi_2$-random variables and present some of their basic properties. We follow closely reference [79], although we prefer to distinguish these two classes of random variables.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a given probability space and denote by $\mathcal{Z} := \mathcal{Z}(\Omega, \mathcal{F}, \mathbb{P})$ the set of (real) random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$. For $1 \leq p < +\infty$, the linear space $L_p := L_p(\Omega, \mathcal{F}, \mathbb{P})$ is the set of random variables $Z$ that satisfies $\mathbb{E}|Z|^p < +\infty$. Identifying two random variables that are equal almost surely (with respect to $\mathbb{P}$), the following function

$$\|Z\|_p := \left(\mathbb{E}|Z|^p\right)^{1/p}. \tag{2.4.1}$$

is a norm on the space $L_p$ that is known as the $L_p$-norm.

Let us consider the function $\|\cdot\|_{\psi_2} : \mathcal{Z} \to \mathbb{R} \cup \{+\infty\}$ given by

$$\|Z\|_{\psi_2} := \sup_{p \geq 1} \frac{\|Z\|_p}{\sqrt{p}}. \tag{2.4.2}$$

We define the set of $\psi_2$-random variables as

$$\psi_2 := \psi_2(\Omega, \mathcal{F}, \mathbb{P}) := \{Z \in \mathcal{Z} : \|Z\|_{\psi_2} < +\infty\}. \tag{2.4.3}$$

Identifying two random variables that are equal almost surely, $\|\cdot\|_{\psi_2}$ is a norm on $\psi_2$, as the notation suggests.

Now, let us consider the definition of a sub-Gaussian random variable.

**Definition 2.4.1.** *We say that $Z \in \mathcal{Z}$ is a $\sigma$-sub-Gaussian random variable, where $\sigma \in [0,+\infty)$, if*

$$M_Z(s) = \mathbb{E}\left[\exp\{sZ\}\right] \leq \exp\{\sigma^2 s^2/2\}, \ \forall s \in \mathbb{R}. \tag{2.4.4}$$

*The constant*

$$\sigma(Z) := \inf\{\sigma \geq 0 : M_Z(s) \leq \exp\{\sigma^2 s^2/2\},\ \forall s \in \mathbb{R}\} \tag{2.4.5}$$

*is the sub-Gaussian moment of* $Z$*.*

The classes of sub-Gaussian and $\psi_2$-random variables are closely related. Let us begin by pointing that every sub-Gaussian random variable is centered. In fact, [54, Proposition 2.1] shows that $\mathbb{E}Z = 0$ if $Z$ is sub-Gaussian. Additionally, it is possible to show that every sub-Gaussian random variable is a $\psi_2$-random variable; see Proposition 2.4.2 below. Reciprocally, if $Z$ is a centered $\psi_2$-random variable, then $Z$ is a sub-Gaussian random variable. So, the space of sub-Gaussian random variables (defined on $(\Omega, \mathcal{F}, \mathbb{P})$) is a linear subspace of $\psi_2$. Observe that a $\psi_2$ random variable need not to be centered (e.g., take $Z = 1$).

In [79] the author does not distinguish these two classes of random variables. Here, we prefer to follow the *classical* definition of sub-Gaussians that uses the moment generating function condition instead of the moment growth condition. To the best of our knowledge, sub-Gaussian random variables were first defined in [34] using its moment generating function and this definition is still usual in the literature nowadays. Of course this is just a matter of taste.

The following proposition, see [79, Lemma 5.5], establishes equivalent properties of $\psi_2$-random variables and relates this class with the sub-Gaussians random variables.

**Proposition 2.4.2.** *Let $Z$ be a random variable. Then the following properties are equivalent, with parameters $K_i > 0$ differing from each other by at most an absolute constant factor.*

1. *Tails:* $\mathbb{P}\left[|Z| \geq s\right] \leq \exp\left(1 - s^2/K_1^2\right),\ \forall s \geq 0$*;*
2. *Moments:* $\|Z\|_p \leq K_2\sqrt{p},\ \forall p \geq 1$*;*
3. *Super-exponential moment:* $\mathbb{E}\left[\exp\left(Z^2/K_3^2\right)\right] \leq e$*.*

*Moreover, if $\mathbb{E}Z = 0$ then properties 1–3 are also equivalent to the following one:*

4. *Moment generating function:* $\mathbb{E}\exp(sZ) \leq \exp(s^2K_4^2)$*, for all* $s \in \mathbb{R}$*.*

*Proof.* See [79, Lemma 5.5]. □

So, the previous proposition establishes that, for each $1 \leq i, j \leq 4$, there exists an absolute constant $C_{i,j}$ such that if $Z$ satisfies property $i$ with constant $K_i$, then $Z$ satisfies property $j$ with constant $C_{i,j}K_i$. Moreover, the additional condition $\mathbb{E}Z = 0$ is necessary for obtaining property (4.) from the remaining ones, although it is not necessary to assume this condition explicitly for the converse statement. In the following corollary, we estimate the constants values $C_{1,2}$, $C_{2,3}$, $C_{3,1}$, $C_{3,4}$ and $C_{4,1}$.

**Corollary 2.4.3.** *We can take the absolute constants of the previous proposition as*

- $C_{1,2} = \sqrt{\frac{e}{2}}$;
- $C_{2,3} = \sqrt{\frac{2}{e-1}}$;
- $C_{3,1} = 1$;
- $C_{3,4} = \sqrt{\frac{e+1}{2}}$;
- $C_{4,1} = 2$.

*Proof.* That we can take $C_{3,1} = 1$ and $C_{4,1} = 2$ can be seen directly in the proof of [79, Lemma 5.5]. Following this proof, it was shown that if $Z$ satisfies (1.) with $K_1 = 1$, then

$$\|Z\|_p \leq \left[\left(\frac{ep}{2}\right)\left(\frac{p}{2}\right)^{p/2}\right]^{1/p} = \frac{1}{\sqrt{2}}\left(\frac{ep}{2}\right)^{1/p}\sqrt{p}. \tag{2.4.6}$$

Now, it is straightforward to verify that $p = 2$ maximizes $p \mapsto \left(\frac{ep}{2}\right)^{1/p}$ on $p \geq 1$. So, we obtain that

$$\frac{\|Z\|_p}{\sqrt{p}} \leq \sqrt{\frac{e}{2}}, \tag{2.4.7}$$

which proves that we can take $C_{1,2}$ as above. In the same proof, it was also shown that:

$$\mathbb{E}\exp(cZ^2) \leq 1 + \sum_{p=1}^{\infty}(2c/e)^p = \frac{1}{1-2c/e} = \frac{e}{e-2c}, \tag{2.4.8}$$

for $0 < c < e/2$. Taking $c = (e-1)/2$, we obtain that $\mathbb{E}\exp(cZ^2) \leq e$, i.e. we can take $C_{2,3} = \sqrt{1/c} = \sqrt{2/(e-1)}$.

Finally, for showing the claim about $C_{3,4}$, we follow the proof of [54, Theorem 3.1]. Let us suppose that for $K_3 = 1$, we have $\mathbb{E}\exp(Z^2/K_3^2) = \mathbb{E}\exp(Z^2) \leq e$. Then, it was shown in the proof of the implication (3) $\Rightarrow$ (1) in [54, Theorem 3.1] (for $a = 1/K_3^2 = 1$) that:

$$\mathbb{E}\exp(sZ) \leq 1 + \frac{s^2}{2}e^{s^2/2}\mathbb{E}e^{Z^2} \leq 1 + \frac{es^2}{2}e^{s^2/2} \leq \left(1 + \frac{es^2}{2}\right)e^{s^2/2} \leq \exp\left\{\frac{s^2}{2}(e+1)\right\}, \tag{2.4.9}$$

for all $s \in \mathbb{R}$. So, we conclude that (4.) is satisfied with $K_4 = \sqrt{\frac{e+1}{2}}$, i.e. $C_{3,4} = \sqrt{\frac{e+1}{2}}$. □

The following result will be useful later.

**Lemma 2.4.4.** *Let $Z$ be a $\psi_2$-random variable. Then,*

$$\|\mathbb{E}Z\|_{\psi_2} = |\mathbb{E}Z| \leq \|Z\|_{\psi_2}. \tag{2.4.10}$$

*Proof.* It is immediate to verify that $\|c\|_{\psi_2} = |c|$, for arbitrary $c \in \mathbb{R}$. Since $Z$ is a $\psi_2$-random variable, we conclude that $\mathbb{E}Z$ is finite, so $\|\mathbb{E}Z\|_{\psi_2} = |\mathbb{E}Z|$. Taking $p = 1$, we obtain:

$$\|Z\|_{\psi_2} \geq \|Z\|_1 / \sqrt{1} = \mathbb{E}\,|Z| \geq |\mathbb{E}Z|\,, \tag{2.4.11}$$

which proves the lemma. □

Now, we present the key result of this section concerning our work. It shows that a $L$-Lipschtiz-transformation of a $\sigma$-sub-Gaussian random variable, after centering, is a $(\kappa L\sigma)$-sub-Gaussian random variable, where $\kappa$ is an absolute constant. Applying Corollary 2.4.3, we estimate the constant $\kappa$[25].

**Proposition 2.4.5.** *Let $Y$ be a random variable with finite expected value such that "$Y - \mathbb{E}Y$" is $\sigma$-sub-Gaussian. If $\psi : \mathbb{R} \to \mathbb{R}$ is a $L$-Lipschitz-continuous function, then $\psi(Y)$ has finite expected value and the random variable $Z := \psi(Y) - \mathbb{E}\psi(Y)$ is $(\kappa L\sigma)$-sub-Gaussian, where*

$$\kappa = 2(C_{1,2}\, C_{2,3}\, C_{3,4}\, C_{4,1}) = 4\sqrt{\frac{e(e+1)}{2(e-1)}} \leq 6.86. \tag{2.4.12}$$

*Proof.* First of all, since $\psi$ is $L$-Lipschitz, we have that $|\psi(Y) - \psi(0)| \leq L|Y - 0| = L|Y|$, so

$$|\psi(Y)| \leq |\psi(0)| + |\psi(Y) - \psi(0)| \leq |\psi(0)| + L|Y|, \tag{2.4.13}$$

which shows that $\mathbb{E}\psi(Y)$ is finite and $Z$ is well-defined. Now, consider the random variable $W := \psi(Y) - \psi(\mathbb{E}Y)$. We have that

$$|W| = |\psi(Y) - \psi(\mathbb{E}Y)| \leq L|Y - \mathbb{E}Y|, \tag{2.4.14}$$

so $\|W\|_p \leq L\,\|Y - \mathbb{E}Y\|_p$ for all $p \geq 1$. This shows that

$$\|W\|_{\psi_2} \leq L\,\|Y - \mathbb{E}Y\|_{\psi_2}\,. \tag{2.4.15}$$

Since $Y - \mathbb{E}Y$ is $\sigma$-sub-Gaussian, it satisfies item (4.) of Proposition 2.4.2 with $K_4 = \sigma/\sqrt{2}$. So, we obtain that:

$$\|Y - \mathbb{E}Y\|_{\psi_2} \leq \frac{C_{4,1}C_{1,2}}{\sqrt{2}}\sigma. \tag{2.4.16}$$

Now, observe that $Z = W - \mathbb{E}W$, so

$$\|Z\|_{\psi_2} \leq \|W\|_{\psi_2} + \|\mathbb{E}W\|_{\psi_2} \leq 2\,\|W\|_{\psi_2} \leq \sqrt{2}(C_{1,2}\, C_{4,1})L\sigma, \tag{2.4.17}$$

where the first inequality is just the triangular inequality and the second one follows from Lemma 2.4.4. Since $\mathbb{E}Z = 0$, applying again Proposition 2.4.2, we obtain that $Z$ satisfies item (4.) with $K_4 = \sqrt{2}(C_{1,2}\, C_{2,3}\, C_{3,4}\, C_{4,1})L\sigma$. So, we conclude that $Z$ is $(\kappa L\sigma)$-sub-Gaussian, with $\kappa = 2(C_{1,2}\, C_{2,3}\, C_{3,4}\, C_{4,1}) = 4\sqrt{\frac{e(e+1)}{2(e-1)}}$. □

[25] Of course, it is possible that the result holds for a smaller $\kappa$.

## 2.5 Convex analysis

In this section we present basic definitions and results of convex analysis that are used throughout the thesis. Let us begin by recalling some basic definitions.

**Definition 2.5.1.** *(extended real numbers) Appending $-\infty$ and $+\infty$ to the set of real numbers, we consider the set of extended real numbers $\overline{R} := \mathbb{R} \cup \{-\infty, +\infty\}$. The appended elements satisfy:*

$$-\infty < x < +\infty, \tag{2.5.1}$$

*for all $x \in \mathbb{R}$. Moreover, it is straightforward to extend the sum and product operations to $\overline{\mathbb{R}}$, except for the following combinations that we consider below:*

$$(-\infty) + (+\infty) = (+\infty) + (-\infty) = +\infty, \tag{2.5.2}$$

$$0 \cdot (\pm\infty) = (\pm\infty) \cdot 0 = 0. \tag{2.5.3}$$

We also assume the following conventions:

$$\sup \emptyset = -\infty; \tag{2.5.4}$$

$$\inf \emptyset = +\infty. \tag{2.5.5}$$

**Definition 2.5.2.** *(convex set) We say that $X \subseteq \mathbb{R}^n$ is convex, if*

$$\lambda x_1 + (1-\lambda)x_2 \in X, \tag{2.5.6}$$

*for all $x_1, x_2 \in X$ and $0 \leq \lambda \leq 1$.*

**Definition 2.5.3.** *(face of a convex set) Let $C$ be a (nonempty) convex set. We say that $F \subseteq C$ is a face of $C$ if $F$ is nonempty and if $F$ satisfies the following property:*

- *if $x, y \in C$ are such that $\lambda x + (1-\lambda)y \in F$, for all $0 < \lambda < 1$, then $x, y \in F$.*

Note that a nonempty convex set is always a face of itself. We say that $F \subseteq C$ is a proper face of $C$, if $F \neq C$ is a face of $C$.

A polyhedral set is an important example of a convex set. This class of sets are very important for the theory of linear programming.

**Definition 2.5.4.** *(polyhedron or polyhedral set) We say that $X \subseteq \mathbb{R}^n$ is a polyhedron or a polyhedral set if there exist a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$ such that*

$$X = \{x \in \mathbb{R}^n : Ax \leq b\}.$$

Note that Definition 2.5.4 implies that every polyhedron is not just convex, but also closed[26].

Now we present the definition of a cone in $\mathbb{R}^n$.

[26] Some authors consider less restrictive definitions for polyhedral sets, but we adopt this one in the thesis.

**Definition 2.5.5.** *(cones) We say that $K \subseteq \mathbb{R}^n$ is a cone if $K \neq \emptyset$ and for every $\lambda \geq 0$ and $x \in K$, we have that $\lambda x \in K$.*

It is elementary to verify that if $K$ is a cone, then $0 \in K$ and $-K := \{-x : x \in K\}$ is also a cone. Cones are commonly used for introducing binary relations in $\mathbb{R}^n$ by the formula

$$x \succeq_K y \Leftrightarrow x - y \in K, \tag{2.5.7}$$

for all $x, y \in \mathbb{R}^n$. As an example consider the positive orthant $\mathbb{R}^n_+ := \{x \in \mathbb{R} : x_i \geq 0, i = 1, \ldots, n\}$. Note that $\mathbb{R}^n_+$ is a cone. It defines the relation $\succeq_{\mathbb{R}^n_+}$ that is a partial ordering in $\mathbb{R}^n$. We have that $x \succeq_{\mathbb{R}^n_+} y$ if and only if $x_i \geq y_i$, for every $i = 1, \ldots, n$. Note also that a polyhedral set can be written as

$$X := \{x \in \mathbb{R}^n : L(x) \in \mathbb{R}^m_+\}, \tag{2.5.8}$$

where $L(x) := b - Ax$ is an affine function. In many interesting examples of optimization problems, it is typical to consider its feasible set on the form

$$X := \{x \in \mathbb{R}^n : G(x) \in K\}, \tag{2.5.9}$$

where $K$ is a closed convex cone and $G : \mathbb{R}^n \to \mathbb{R}^m$ is a function.

**Definition 2.5.6.** *(pointed cones) We say that $K \subseteq \mathbb{R}^n$ is a pointed cone if $K$ is a cone and whenever $x, -x \in K$, we must have that $x = 0$.*

**Definition 2.5.7.** *(Properties of a binary relation in $\mathbb{R}^n$) Let $\succeq \subseteq \mathbb{R}^n \times \mathbb{R}^n$ be a binary relation. Consider the following properties:*

*(i) Reflexivity: for every $x \in \mathbb{R}^n$, $x \succeq x$.*

*(ii) Transitivity: for every $x, y, z \in \mathbb{R}^n$, if $x \succeq y$ and $y \succeq z$, then $x \succeq z$.*

*(iii) Anti-symmetry: for every $x, y \in \mathbb{R}^n$, if $x \succeq y$ and $y \succeq x$, then $x = y$.*

*(iv) Homogeneity: for every $x, y \in \mathbb{R}^n$ and $\lambda \geq 0$, if $x \succeq y$, then $\lambda x \geq \lambda y$.*

*(v) Additivity: for every $w, x, y, z \in \mathbb{R}^n$, if $x \succeq y$ and $w \succeq z$, then $x + w \succeq y + z$.*

*(vi) Continuity: for every sequences $\{x^j : j \in \mathbb{N}\}$ and $\{y^j : j \in \mathbb{N}\}$ in $\mathbb{R}^n$, if $x^j \succeq y^j$, for every $j \in \mathbb{N}$, $x^j \to x$ and $y^j \to y$, then $x \succeq y$.*

The following proposition relates properties of the cone $K$ with properties of the binary relation $\succeq_K$ induced by $K$.

**Proposition 2.5.8.** *Take any cone $K \subseteq \mathbb{R}^n$. The following assertions hold:*

*(i) $\succeq_K$ is reflexive and homogeneous.*

*(ii) If $K$ is convex, then $\succeq_K$ is also transitive and satisfies the additivity property.*

*(iii) If $K$ is pointed, then $\succeq_K$ is anti-symmetric.*

*(iv) If $K$ is closed, then $\succeq_K$ is continuous.*

*Proof.* (i) Since $x - x = 0 \in K$, it follows that $x \succeq_K x$. Moreover, if $x \succeq_K y$ and $\lambda \geq 0$, then $\lambda x - \lambda y = \lambda(x - y) \in K$.

(ii) Note that if $K$ is a convex cone, then $K + K = K$. Since $0 \in K$, the inclusion $\supseteq$ is trivial. Take any $x, y \in K$. Note that $x + y = 2(x/2 + y/2) \in K$, which proves the converse inclusion. The transitivity and the additivity properties follow easily from this equality $K + K = K$.

(iii) Let $K$ be a pointed cone. If $x \succeq_K y$ and $y \succeq_K x$, then $x - y \in K$ and $y - x = -(x - y) \in K$. It follows that $x - y = 0$, i.e. $x = y$.

(iv) Note that $x^j - y^j \in K$, for every $j \in \mathbb{N}$. Moreover, $x - y = \lim_j (x^j - y^j) \in \overline{K} = K$, i.e. $x \succeq_K y$. □

Let us recall that a binary relation $\succeq$ in $\mathbb{R}^n$ is said to be a partial order if it is reflexive, transitive and anti-symmetrical. It follows that if $K$ is a pointed convex cone, then $\succeq_K$ is a partial order in $K$. The following lemma will be useful.

**Lemma 2.5.9.** *Take any convex cone $K \subseteq \mathbb{R}^n$. If $x \in K$ and $y \in \operatorname{int} K$, then $x + y \in \operatorname{int} K$.*

*Proof.* Since $y \in \operatorname{int} K$, there exists $\epsilon > 0$ such that $B(y; \epsilon) \subseteq K$. By item (ii)[27] of Proposition 2.5.8 it follows that

$$B(x + y; \epsilon) = x + B(y; \epsilon) \subseteq K + K = K, \tag{2.5.10}$$

i.e. $x + y \in \operatorname{int} K$. □

**Definition 2.5.10.** *(affine set) We say that $A \subseteq \mathbb{R}^n$ is an affine set if for every $x_1, x_2 \in A$, we have that:*

$$(1 - t)x_1 + tx_2 \in A, \ \forall t \in \mathbb{R}.$$

**Definition 2.5.11.** *(affine hull of a set) Let $X \subset \mathbb{R}^n$ be given. We define the affine hull of $X$ as:*

$$\operatorname{aff} X := \bigcap \{A : A \supseteq X, A \textit{ is an affine set } \}.$$

It can be shown that an arbitrary intersection of affine sets is also an affine set. It follows that the affine hull of $X$ is the smallest affine set that contains $X$. This notion is crucial to define the relative interior of a set. We will recall interesting properties of the the relative interior of convex sets.

[27] We have shown in the proof of Proposition 2.5.8 that if $K$ is a convex cone, then $K + K = K$.

**Definition 2.5.12.** *(relative interior of a set) Let $X \subseteq \mathbb{R}^n$ be a nonempty set. We say that $x \in \operatorname{ri} X$ if there exists $\epsilon > 0$ such that:*

$$\mathbb{B}(x;\epsilon) \cap \operatorname{aff} X \subseteq X. \tag{2.5.11}$$

*Let us define* $\operatorname{ri} \emptyset := \emptyset$.

When $X$ is nonempty convex its relative interior is nonempty[28]. In fact, the following result holds:

**Proposition 2.5.13.** *Let $X \subseteq \mathbb{R}^n$ be a convex set. Then, $\overline{\operatorname{ri} X} = \overline{X}$.*

*Proof.* See [7, Proposition 1.3.5]. □

**Definition 2.5.14.** *(epigraph of a real-valued function) Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function. The epigraph of $f$ is the set:*

$$\operatorname{epi} f := \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq t\}. \tag{2.5.12}$$

**Definition 2.5.15.** *(domain of a real-valued function) Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function. The domain of $f$ is the set:*

$$\operatorname{dom} f := \{x \in \mathbb{R}^n : f(x) < +\infty\}. \tag{2.5.13}$$

**Definition 2.5.16.** *(proper functions) We say that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper if $\operatorname{dom} f \neq \emptyset$ and $f(x) > -\infty$, for all $x \in \mathbb{R}$.*

**Definition 2.5.17.** *(convex functions) We say that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a convex function if $\operatorname{epi} f \subseteq \mathbb{R}^{n+1}$ is a convex set.*

We can also consider the definition of a vector-valued convex function with respect to a binary relation $\succeq$ in $\mathbb{R}^m$.

**Definition 2.5.18.** *Take any binary relation $\succeq$ in $\mathbb{R}^m$. We say that $G : \mathbb{R}^n \to \mathbb{R}^m$ is convex with respect to $\succeq$ if*

$$G(\lambda x + (1-\lambda)y) \preceq \lambda G(x) + (1-\lambda)G(y), \tag{2.5.14}$$

*for every $x, y \in \mathbb{R}^n$ and for every $0 \leq \lambda \leq 1$.*

**Definition 2.5.19.** *(polyhedral functions) We say that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a polyhedral function if it is proper and $\operatorname{epi} f \subseteq \mathbb{R}^{n+1}$ is a polyhedron.*

Note that every polyhedral function is l.s.c. (see Definition 2.5.24 and Proposition 2.5.25) and convex.

[28] That is not the case with the topological interior. For example, $X = \{0\} \subseteq \mathbb{R}$ is such that $\operatorname{int} X = \emptyset$.

**Definition 2.5.20.** *(subdifferential of a convex function) Let $f : \mathbb{R}^n \to \overline{R}$ be a convex function. We define the subdifferential set of $f$ at $x \in \mathbb{R}^n$ as:*

$$\partial f(x) := \{s \in \mathbb{R}^n : f(y) \geq f(x) + \langle s, y - x\rangle\ , \text{ for all } y \in \mathbb{R}^n\}. \tag{2.5.15}$$

*We denote the elements of $\partial f(x)$ as the subgradients of $f$ at $x$.*

For $x \in \operatorname{dom} f$, we have that:

$$\partial f(x) = \bigcap_{y \in \operatorname{dom} f} \{s \in \mathbb{R}^n : f(y) \geq f(x) + \langle s, y - x\rangle\}. \tag{2.5.16}$$

Since each set in the intersection is closed and convex, it follows that $\partial f(x)$ is also closed and convex (this is trivially true for $x \notin \operatorname{dom} f$). The following proposition gives a sufficient condition for the existence of subgradients of $f$ at a point $x \in \mathbb{R}^n$.

**Proposition 2.5.21.** *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function. For every $x \in \operatorname{ri} dom f$, we have that $\partial f(x) \neq \emptyset$. Moreover, if $x \in \operatorname{int} \operatorname{dom} f$, we conclude (additionally) that $\partial f(x)$ is compact.*

*Proof.* This proposition is an immediate consequence of [7, Prop. 5.4.1]. □

**Definition 2.5.22.** *(convex conjugate of real-valued functions) Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be given. The convex conjugate of $f$ is:*

$$f^*(y) := \sup_{x \in \mathbb{R}^n} \{\langle y, x\rangle - f(x)\} = \sup_{x \in \operatorname{dom} f} \{\langle y, x\rangle - f(x)\}. \tag{2.5.17}$$

The following result gives a characterization of the subgradients of a convex function $f$ at $x$ in terms of the values assumed by $f$ and $f^*$.

**Proposition 2.5.23.** *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function. We have that:*

$$s \in \partial f(x) \Leftrightarrow \langle s, x\rangle = f(x) + f^*(s). \tag{2.5.18}$$

*Proof.* See [7, Proposition 5.4.3]. Note that the result is trivially true if $f$ is not proper. Indeed, in that case, none of the conditions can hold. □

**Definition 2.5.24.** *(lower semi-continuous functions) We say that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is lower semi-continuous (l.s.c.) at $x \in \mathbb{R}^n$ if for all $t \in \mathbb{R}$ such that $f(x) > t$, there exists an $\epsilon > 0$ such that $f(y) > t$, for all $y \in \mathbb{B}(x; \epsilon)$. We say that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is l.s.c. if $f$ is l.s.c. at every $x \in \mathbb{R}^n$.*

The following result is well-known.

**Proposition 2.5.25.** *A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is l.s.c. if and only if $\operatorname{epi} f \subseteq \mathbb{R}^{n+1}$ is closed.*

*Proof.* See [7, Proposition 1.1.2]. □

**Definition 2.5.26.** *(closure of a function) Let $f : \mathbb{R}^n \to \overline{R}$ be a function. We define the closure of $f$ as the unique function* $\operatorname{cl} f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *that satisfies*[29]

$$\operatorname{epi}(\operatorname{cl} f) = \overline{\operatorname{epi} f}.$$

Of course, $\operatorname{cl} f \leq f$. By Proposition 2.5.25, we have that $f$ is l.s.c. if and only if $\operatorname{cl} f = f$. The function $\operatorname{cl} f$ is the greater l.s.c. function that is below $f$. Finally, if $f$ is convex, then $\operatorname{cl} f$ is also convex.

**Proposition 2.5.27.** *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be arbitrary. We have that $f^*$ is l.s.c. and convex.*

*Proof.* We just need to show that $\operatorname{epi}(f^*)$ is closed and convex. Observe that $f^*$ is the supremum of the affine functions:

$$L_x(s) := \langle x, s \rangle - f(x),$$

that are, in particular, l.s.c. and convex. It follows that $\operatorname{epi} L_x$ is closed and convex, for every $x \in \mathbb{R}^n$. So,

$$\operatorname{epi}(f^*) = \bigcap_{x \in \mathbb{R}^n} \operatorname{epi} L_x$$

is closed and convex. □

Let us recall this classic result of convex analysis.

**Theorem 2.5.28.** *(Fenchel-Moreau theorem) Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper l.s.c. convex function. Then,*

$$f^{**} = f. \tag{2.5.19}$$

*Proof.* See [7, Proposition 1.6.1] or [59, Pag. 474, Theorem 11.1]. □

Now, we introduce the definition of the directional derivative of a function.

**Definition 2.5.29.** *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a function. We define the directional derivative of $f$ at $x \in \operatorname{dom} f$ on direction $d \in \mathbb{R}^n$ as the limit (if it exists!):*

$$f'(x; d) := \lim_{t \to 0+} \left\{ \frac{f(x + td) - f(x)}{t} \right\}. \tag{2.5.20}$$

When $f$ is convex, the limit (2.5.20) exists, possibly assuming the values $\pm\infty$, for every $x \in \operatorname{dom} f$ (see, for example, [7, Page 196]).

The following variational characterization of the directional derivative $f'(x; \cdot)$ of a convex function $f$ at a point $x \in \operatorname{dom} f$ in terms of the subdifferential of $f$ at $x$ is particularly useful.

[29] In fact, it is possible to show that if $E \subseteq \mathbb{R}^{n+1}$ is the epigraph of a function, then $\overline{E}$ is also the epigraph of a function. For instance, see [7, Section 1.3.3]

**Proposition 2.5.30.** *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper convex function and $x \in \operatorname{dom} f$. The following conditions hold:*

*(i) When $\partial f(x) \neq \emptyset$, we have that:*

$$\operatorname{cl} f'(x; \cdot) = \sup_{s \in \partial f(x)} \langle s, \cdot \rangle . \tag{2.5.21}$$

*(ii) If $x \in \operatorname{ri} \operatorname{dom} f$, we have that:*

$$f'(x; \cdot) = \sup_{s \in \partial f(x)} \langle s, \cdot \rangle . \tag{2.5.22}$$

*Proof.* See [7, Proposition 5.4.8]. □

Now, let us consider convex functions $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ defined on the set of real numbers. For $x < y < z$, $y \in \operatorname{dom} f$, we have the secant inequality (see [7, Page 196], for $d = \pm 1$):

$$\frac{f(y) - f(x)}{y - x} \leq \frac{f(z) - f(y)}{z - y}. \tag{2.5.23}$$

It is also clear from Definition (2.5.20) that, for $x \in \operatorname{dom} f$, $f'(x; 1)$ and $-f'(x; -1)$ are the right and left derivatives of $f$ at $x$, respectively. Considering also inequality (2.5.23), we have that:

$$f'(x; 1) = \lim_{y \to x+} \frac{f(y) - f(x)}{y - x} = \inf_{y > x} \frac{f(y) - f(x)}{y - x}, \tag{2.5.24}$$

$$-f'(x; -1) = \lim_{y \to x-} \frac{f(y) - f(x)}{y - x} = \sup_{y < x} \frac{f(y) - f(x)}{y - x}. \tag{2.5.25}$$

For $x < y < z$, $y \in \operatorname{dom} f$, we can bound the left and right derivatives of $f$ at $y$ by the slopes:

$$\frac{f(y) - f(x)}{y - x} \leq -f'(y; -1) \leq f'(y; 1) \leq \frac{f(z) - f(y)}{z - y}. \tag{2.5.26}$$

When a proper convex function $f$ is defined on $\mathbb{R}$, the directional derivative $f'(x; \cdot)$ is l.s.c., for every $x \in \mathbb{R}$ such that $\partial f(x) \neq \emptyset$. Let us point out that this is not necessarily the case when $f$ is defined on $\mathbb{R}^d$, for $d > 1$.

**Lemma 2.5.31.** *Let $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be a proper convex function and $x \in \operatorname{dom} f$ be given. If $\partial f(x) \neq \emptyset$, then $f'(x; \cdot)$ is a l.s.c. function, i.e. $\operatorname{cl} f'(x; \cdot) = f'(x; \cdot)$.*

*Proof.* Let $x \in \operatorname{dom} f$ be arbitrary. It is immediate to verify that $f'(x; \cdot)$ is positively homogeneous. In fact, let $s > 0$ be given, we have that:

$$f'(x; sd) = \lim_{t \to 0+} \frac{f(x + t(sd)) - f(x)}{t} = s \lim_{t \to 0+} \frac{f(x + (ts)d)) - f(x)}{ts} = sf'(x; d).$$

So, we conclude that:

$$f'(x;s) = \begin{cases} sf'(x;1), & \text{when } s > 0; \\ 0, & \text{when } s = 0; \\ sf'(x;-1), & \text{when } s < 0. \end{cases} \tag{2.5.27}$$

From the equality above, we obtain that whenever $f'(x;1) \in (-\infty, +\infty] \ni f'(x;-1)$, $f'(x;\cdot)$ is a l.s.c. function. For proving the lemma, we only need to show that $f'(x;1) > -\infty$ and $f'(x;-1) > -\infty$. By hypothesis, there exists $s \in \partial f(x)$. So, for $t > 0$, we have that:

$$\frac{f(x+t) - f(x)}{t} \geq s > -\infty, \tag{2.5.28}$$

and letting $t \to 0+$, we obtain that $f'(x;1) \geq s > -\infty$. Finally, for $t > 0$, we also have that:

$$f(x-t) - f(x) \geq -st, \tag{2.5.29}$$

that is:

$$\frac{f(x-t) - f(x)}{t} \geq -s > -\infty, \tag{2.5.30}$$

and letting $t \to 0+$, we obtain that $f'(x;-1) \geq -s > -\infty$. □

We introduce next some notation about intervals of $\overline{\mathbb{R}}$.

**Remark 2.5.32.** *(intervals on $\overline{\mathbb{R}}$) Let $a$ and $b$ be elements of $\overline{\mathbb{R}}$. We define the interval:*

$$[a,b] := \{x \in \overline{\mathbb{R}} : a \leq x \text{ and } x \leq b\}. \tag{2.5.31}$$

*Observe that we do not assume, in principle, that $a \leq b$. So, if $a > b$, we have that $[a,b] = \emptyset$. One way to restrict the interval $[a,b]$ to the set of real numbers is to consider $[a,b] \cap \mathbb{R}$. Of course, when $a, b \in \mathbb{R}$, we have that $[a,b] = [a,b] \cap \mathbb{R}$. Observe that it is not true in general that:*

$$\inf[a,b] = a, \text{ or } \sup[a,b] = b. \tag{2.5.32}$$

*Consider, for example, $a = 1$ and $b = 0$. We have that $[a,b] = \emptyset$, $\inf[a,b] = +\infty > a$ and $\sup[a,b] = -\infty < b$. This is a somewhat pathological situation. In fact, the following conditions rule out this situation:*

*(i) if $[a,b] \neq \emptyset$, then $\inf[a,b] = a$ and $\sup[a,b] = b$.*

*(ii) if $[a,b] \cap \mathbb{R} \neq \emptyset$, then $\inf([a,b] \cap \mathbb{R}) = a$ and $\sup([a,b] \cap \mathbb{R}) = b$.*

The following result is a corollary of the previous proposition, lemma and remark.

**Corollary 2.5.33.** *Let be given a proper convex function $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ and $x \in \operatorname{dom} f$. We have that:*

$$\partial f(x) = [-f'(x;-1), f'(x;1)] \cap \mathbb{R}. \tag{2.5.33}$$

*Moreover, we also have that:*

$$\begin{aligned} \inf \partial f(x) &= -f'(x;-1) \text{ and} && (2.5.34) \\ \sup \partial f(x) &= f'(x;1), && (2.5.35) \end{aligned}$$

*when $\partial f(x) \neq \emptyset$.*

*Proof.* Let $x \in \operatorname{dom} f$ be given. First of all, let us show that:

$$\partial f(x) \subseteq [-f'(x;-1), f'(x;1)] \cap \mathbb{R}.$$

If $\partial f(x) = \emptyset$, then there is nothing to prove. Suppose that $\partial f(x) \neq \emptyset$. By Proposition 2.5.30 and by Lemma 2.5.31, we have that:

$$\begin{aligned} f'(x;1) = \operatorname{cl} f'(x;1) &= \sup_{s \in \partial f(x)} s = \sup \partial f(x), && (2.5.36) \\ -f'(x;-1) = -\operatorname{cl} f'(x;-1) &= -\sup_{s \in \partial f(x)} -s = -(-\inf \partial f(x)) = \inf \partial f(x). && (2.5.37) \end{aligned}$$

So, for $s \in \partial f(x)$, we have that $\inf \partial f(x) \leq s \leq \sup \partial f(x)$ and $s \in \mathbb{R}$, which proves the inclusion $\partial f(x) \subseteq [-f'(x;-1), f'(x;1)] \cap \mathbb{R}$.

Now let us show the converse inclusion $[-f'(x;-1), f'(x;1)] \cap \mathbb{R} \subseteq \partial f(x)$. Again, if $[-f'(x;-1), f'(x;1)] \cap \mathbb{R} = \emptyset$, then there is nothing to be proved. Suppose that $[-f'(x;-1), f'(x;1)] \cap \mathbb{R} \neq \emptyset$ and take $s \in [-f'(x;-1),\ f'(x;1)] \cap \mathbb{R}$. For $y > x$, we get from equation (2.5.23) that:

$$\frac{f(y) - f(x)}{y - x} \geq f'(x;1) \geq s. \tag{2.5.38}$$

For $y < x$, we get again from equation (2.5.23) that:

$$\frac{f(y) - f(x)}{y - x} \leq -f'(x;-1) \leq s. \tag{2.5.39}$$

We conclude that $[-f'(x;-1), f'(x;1)] \cap \mathbb{R} \subseteq \partial f(x)$. Finally, suppose that $\partial f(x) \neq \emptyset$. By Remark 2.5.32, we obtain that:

$$\begin{aligned} \inf \partial f(x) &= -f'(x;-1), \text{ and} && (2.5.40) \\ \sup \partial f(x) &= f'(x;1). && (2.5.41) \end{aligned}$$

This concludes the proof of the corollary. □

Of course when $\partial f(x) \neq \emptyset$ is bounded, we have that:

$$\partial f(x) = \left[-f'(x;-1), f'(x;1)\right].$$

Now, let us consider a basic lemma from real analysis.

**Lemma 2.5.34.** *Let $\{I(s) : s \in J\}$ be a family of nonempty closed intervals on $\mathbb{R}$, where $J \subseteq \mathbb{R}$ is also an interval. Let $a(s) \leq b(s)$ be the extreme points of $I(s)$, for all $s \in J$. Suppose that the following conditions are satisfied:*

*(i) For every $s < t$ on $J$, we have that $b(s) \leq a(t)$.*

*(ii) For every $s \in J$ such that $s < \sup J$:*

$$b(s) = \lim_{\substack{t \to s+ \\ t \in J}} a(t);$$

*(iii) For every $s \in J$ such that $s > \inf J$:*

$$a(s) = \lim_{\substack{t \to s- \\ t \in J}} b(t).$$

*Then, it follows that:*

$$I := \bigcup_{s \in J} I(s)$$

*is an interval on $\mathbb{R}$.*

*Proof.* If $I$ is empty or a singleton, then there is nothing to prove. So, let $y_1 < y_2$ be given elements of $I$. We just need to show that $(y_1, y_2) \subseteq I$. For such, let $y_1 < y < y_2$ be arbitrary. First of all, observe that there exist $s_1, s_2 \in J$ such that $y_1 \in I(s_1)$ and $y_2 \in I(s_2)$; that is:

$$a(s_1) \leq y_1 < y < y_2 \leq b(s_2).$$

We will show that there exists $t \in J$ such that $y \in I(t)$. Define the following sets:

$$J_- := \{s \in J : b(s) < y\}; \tag{2.5.42}$$

$$J_+ := \{s \in J : a(s) > y\}. \tag{2.5.43}$$

Note that if $J_- = \emptyset$ or $J_+ = \emptyset$, we obtain that $y \in I(s_1)$ or $y \in I(s_2)$, respectively. In fact, if $J_- = \emptyset$, then $s_1 \notin J_-$, i.e. $y \leq b(s_1)$ and it follows that $y \in I(s_1) \subseteq I$. Similarly, if $J_+ = \emptyset$, we obtain that $y \in I(s_2) \subseteq I$. So, let us suppose without loss of generality that $J_-$ and $J_+$ are both nonempty. We claim that there exists $t \in J \setminus J_- \cup J_+$. Observe that if such $t$ exists, then:

$$a(t) \leq y \leq b(t),$$

i.e. $y \in I(t) \subseteq I$.

Note that $J_- \cap J_+ = \emptyset$, since if $s \in J_-$, then $a(s) \leq b(s) < y$, so $s \notin J_+$. We also have that $J_-$ and $J_+$ are open on $J$. In fact, let us show the claim for $J_-$[30]. Let $s \in J_-$ be given. Let us show that there exists an $\epsilon > 0$ such that $(s - \epsilon, s + \epsilon) \cap J \subseteq J_-$. Of course, if $s' \in J$ is such that $s' < s$, then: $b(s') \leq a(s) \leq b(s) < y$. i.e. $s' \in J_-$. Suppose by contradiction that for every $\epsilon > 0$, there exists $s < s' < s + \epsilon$ such that $b(s') \geq y$. Then we can construct a decreasing sequence $(s'_k : k \in \mathbb{N})$ on $J$ converging to $s$ such that $b(s'_k) \geq y$, for all $k \in \mathbb{N}$. We have that:

$$a(s'_k) \geq b(s'_{k+1}) \geq y, \ \forall k \in \mathbb{N}.$$

Letting $k \to +\infty$, we obtain that $y \leq \lim_{k \in \mathbb{N}} a(s'_k) = b(s) < y$ ($\to\leftarrow$).

Since $J$ is connected and we are supposing that $J_- \neq \emptyset \neq J_+$, it cannot occur that $J = J_- \cup J_+$, which concludes the proof. □

**Proposition 2.5.35.** *Let $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be a proper convex function. Consider the family of sets $\{\partial f(x) : x \in \mathbb{R}\}$. For every $x \in \mathbb{R}$, $\partial f(x) = \emptyset$ or $\partial f(x)$ is a nonempty closed interval that has as extreme points $-f'(x; -1) \leq f'(x; 1)$. Moreover, $\{x \in \mathbb{R} : \partial f(x) \neq \emptyset\}$ is a nonempty interval. The following conditions are also satisfied:*

*(i) For every $x, y \in \operatorname{dom} f$ such that $x < y$, we have that: $f'(x; 1) \leq -f'(y; -1)$;*

*(ii) For every $x \in \operatorname{dom} f$ satisfying $x < \sup(\operatorname{dom} f)$, we have that:*

$$f'(x;1) = \lim_{\substack{y \to x+ \\ y \in \operatorname{dom} f}} f'(y;1) = \lim_{\substack{y \to x+ \\ y \in \operatorname{dom} f}} -f'(y;-1). \quad (2.5.44)$$

*(iii) For every $x \in \operatorname{dom} f$ satisfying $x > \inf(\operatorname{dom} f)$, we have that:*

$$-f'(x;1) = \lim_{\substack{y \to x- \\ y \in \operatorname{dom} f}} -f'(y;-1) = \lim_{\substack{y \to x- \\ y \in \operatorname{dom} f}} f'(y;1). \quad (2.5.45)$$

*Proof.* Let $x \in \mathbb{R}$ be such that $\partial f(x) \neq \emptyset$. It follows that $x \in \operatorname{dom} f$ and, by Corollary 2.5.33, that $\partial f(x)$ is a (nonempty) closed interval with extreme points $-f'(x; -1) \leq f'(x; 1)$.

[30]The proof of the fact that $J_+$ is open on $J$ is similar.

Now, let $x, y \in \operatorname{dom} f$ satisfying $x < y$ be given. By equation (2.5.23), we have that:

$$f'(x;1) \leq \frac{f(y)-f(x)}{y-x} \leq -f'(y;-1), \tag{2.5.46}$$

which shows item (i). Moreover, since $-f'(z;-1) \leq f'(z;1)$, for all $z \in \operatorname{dom} f$, it follows that both functions $-f'(\cdot;-1)$ and $f'(\cdot;1)$ are monotonically non-decreasing on $\operatorname{dom} f$.

Since $f$ is proper, we have that $\operatorname{dom} f \neq \emptyset$, and so $\operatorname{ri}\operatorname{dom} f \neq \emptyset$. Moreover, $J := \{x \in \mathbb{R} : \partial f(x) \neq \emptyset\} \supseteq \operatorname{ri}\operatorname{dom} f$, by Proposition 2.5.21, and we conclude that $J$ is nonempty. Let us show that $J$ is an interval. Let be given $x < y$ such that $\partial f(x) \neq \emptyset \neq \partial f(y)$. We will show that $(x, y) \subseteq J$. In fact, we have that $x$ and $y$ belong to $\operatorname{dom} f$. By the convexity of $f$, it follows that $(x, y) \subseteq \operatorname{dom} f$. Since $(x, y)$ is open, we also conclude that $(x, y) \subseteq \operatorname{int}\operatorname{dom} f$ and, by Proposition 2.5.21, $\partial f(z) \neq \emptyset$, for all $z \in (x, y)$.

Now, we show item (ii). Let $x \in \operatorname{dom} f$ satisfying $x < \sup(\operatorname{dom} f)$ be given. Since $f'(x;1) \leq -f'(y;-1) \leq f'(y;1)$, for all $x < y \in \operatorname{dom} f$, we conclude that:

$$f'(x;1) \leq \lim_{y\to x+,\, y\in\operatorname{dom} f} -f'(y;-1) \leq \lim_{y\to x+,\, y\in\operatorname{dom} f} f'(y;1),$$

where the limits exist by the monotonicity of $-f'(\cdot;-1)$ and $f'(\cdot;1)$ on $\operatorname{dom} f$. Now we show that the opposite inequalities are also satisfied. Given $\epsilon > 0$ there exists $\bar{y} > x$ ($\bar{y} \in \operatorname{dom} f$) such that:

$$\frac{f(y)-f(x)}{y-x} \leq f'(x;1) + \epsilon/2, \forall x < y \leq \bar{y}. \tag{2.5.47}$$

So, taking $x < z \leq (x+\bar{y})/2$ arbitrary and considering $y := z + (z-x)$, we obtain that: $z < y \leq \bar{y}$ and $(z-x)/(y-x) = 1/2 = (y-z)/(y-x)$. It follows that:

$$\begin{aligned}
f'(x;1) + \epsilon/2 &\geq \frac{f(y)-f(x)}{y-x} && (2.5.48)\\
&= \frac{f(y)-f(z)}{y-z}\frac{y-z}{y-x} + \frac{f(z)-f(x)}{z-x}\frac{z-x}{y-x} && (2.5.49)\\
&\geq \frac{1}{2}f'(z;1) + \frac{1}{2}f'(x;1) && (2.5.50)\\
&\geq \frac{1}{2}\lim_{w\to x+,\, w\in\operatorname{dom} f} f'(w;1) + \frac{1}{2}f'(x;1). && (2.5.51)
\end{aligned}$$

So, for every arbitrary $\epsilon > 0$, we conclude that:

$$f'(x;1) + \epsilon \geq \lim_{y\to x+,\, y\in\operatorname{dom} f} f'(y;1) \geq \lim_{y\to x+,\, y\in\operatorname{dom} f} -f'(y;-1); \tag{2.5.52}$$

i.e. $\lim_{y\to x+,\, y\in\operatorname{dom} f} -f'(y;-1) \leq \lim_{y\to x+,\, y\in\operatorname{dom} f} f'(y;1) \leq f'(x;1)$, which proves item (ii).

The proof of item (iii) is similar to the previous one. Let $x \in \operatorname{dom} f$ satisfying $x > \inf(\operatorname{dom} f)$ be given. It follows immediately that $-f'(x;-1) \geq \lim_{y\to x-,\, y\in\operatorname{dom} f} f'(y;1) \geq \lim_{y\to x-,\, y\in\operatorname{dom} f} -f'(-y;1)$ and that the limits are well-defined. Given $\epsilon > 0$, there exists $\bar{y} < x$ ($\bar{y} \in \operatorname{dom} f$) such that:

$$-f'(x;-1) - \frac{\epsilon}{2} \leq \frac{f(x)-f(y)}{x-y},$$

for all $\bar{y} \leq y < x$. So, taking $(x+\bar{y})/2 \leq z < x$ arbitrary and considering $y := z-(x-z)$, it follows that:

$$\begin{aligned} -f'(x;-1) - \frac{\epsilon}{2} &\leq \frac{f(x)-f(y)}{x-y} && (2.5.53)\\ &= \frac{1}{2}\frac{f(x)-f(z)}{x-z} + \frac{1}{2}\frac{f(z)-f(y)}{z-y} && (2.5.54)\\ &\leq \frac{1}{2}(-f'(x;-1)) + \frac{1}{2}(-f'(z;-1)), && (2.5.55) \end{aligned}$$

which implies that:

$$-f'(x;-1) \leq -f'(z;-1)+\epsilon \leq \lim_{y\to x-,\, y\in\operatorname{dom} f} -f'(y;-1)+\epsilon \leq \lim_{y\to x-,\, y\in\operatorname{dom} f} f'(y;1)+\epsilon. \tag{2.5.56}$$

Since $\epsilon > 0$ is arbitrary, we conclude the proof of item (iii). $\square$

**Corollary 2.5.36.** *Let $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be a proper convex function. We have that:*

$$\bigcup_{x\in dom f} \partial f(x) \tag{2.5.57}$$

*is a nonempty interval on $\mathbb{R}$.*

*Proof.* Immediate from Lemma 2.5.34 and Proposition 2.5.35. $\square$

Now let us consider the definition of the set of all subgradients of a convex function $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$.

**Definition 2.5.37.** *(set of subgradients of a convex function) Let $f : \mathbb{R} \to \mathbb{R}\cup\{+\infty\}$ be a convex function. We define the set of all subgradients of $f$ by:*

$$\mathcal{I}_f := \bigcup_{x\in\mathbb{R}} \partial f(x) = \bigcup_{x\in\operatorname{dom} f} \partial f(x). \tag{2.5.58}$$

*We also denote its extreme points by:*

$$l(f) := \inf \mathcal{I}_f \geq -\infty; \tag{2.5.59}$$

$$L(f) := \sup \mathcal{I}_f \leq +\infty. \tag{2.5.60}$$

When $f$ is proper, we have that $\mathcal{I}_f$ is an nonempty interval and $l(f) \leq L(f)$. When $f$ is not proper, i.e. $f(x) = +\infty$, for all $x \in \mathbb{R}$, we have that $\mathcal{I}_f = \emptyset$ and $l(f) = +\infty > -\infty = L(f)$. In the next proposition we study the extreme points of $\mathcal{I}_\phi$, for $\phi \in \Phi$.

**Proposition 2.5.38.** *Let $\phi \in \Phi$ be given. The following statements hold:*

*(i)* $0 \leq l(\phi) \leq 1$*;*

*(ii)* $1 \leq L(\phi) \leq +\infty$*;*

*(iii)* *$\phi$ is Lipschitz continuous if and only if $L(\phi) < +\infty$. In that case, $L(\phi)$ is the (smallest possible) Lipschitz constant of $\phi$.*

*Proof.* Take $\phi \in \Phi$ and $s \in \mathcal{I}_\phi$. There exists $x \in \operatorname{dom}\phi$ such that $s \in \partial\phi(x)$. Since $\phi$ is monotonically non-decreasing, we conclude that $s \geq 0$. In fact:

$$\phi(x-1) \geq \phi(x) + s(-1),$$

which implies that:

$$s \geq \phi(x) - \phi(x-1) \geq 0.$$

Moreover, we have that $1 \in \partial\phi(0) \subseteq \mathcal{I}_\phi$, so $0 \leq l(\phi) = \inf \mathcal{I}_\phi \leq 1$ and item (i) is proved. Item (ii) follows also immediately from the fact that $1 \in \partial\phi(0)$.

Now, let us prove item (iii). Suppose that $\phi$ is $L$-Lipschitz continuous, where $0 \leq L < +\infty$. Since $\phi$ is proper (in fact, $\phi(0) = 0$) we conclude that $\phi$ is finite-valued. Indeed,

$$|\phi(x)| = |\phi(x) - \phi(0)| \leq L|x|, \ \forall x \in \mathbb{R}.$$

Let $s \in \partial\phi(x)$ be given, where $x \in \mathbb{R}$ is arbitrary. Taking $y > x$ we obtain that:

$$L(y-x) \geq f(y) - f(x) \geq s(y-x),$$

i.e. $s \leq L$. Taking the supremum for $s$ on $\mathcal{I}_\phi$, we conclude that $L(\phi) \leq L < +\infty$.

Reciprocally, suppose that $L(\phi) < +\infty$. First of all, let us show that $\phi$ is $L(\phi)$-Lipschitz continuous on $\operatorname{dom}\phi$. Then, we will show that $\operatorname{dom}\phi = \mathbb{R}$. Let $x < y$ be elements of $\operatorname{dom}\phi$. By Theorem 2.5.40, we have that:

$$|\phi(y) - \phi(x)| = \phi(y) - \phi(x) = \int_x^y \phi'(t;1)dt \leq L(\phi)(y-x), \tag{2.5.61}$$

which proves that $\phi$ is $L(\phi)$-Lipschitz on $\operatorname{dom}\phi$. Taking $x = 0$ in equation (2.5.61), we obtain that:

$$\phi(y) \leq L(\phi)y, \ \forall 0 < y \in \operatorname{dom}\phi. \tag{2.5.62}$$

Since $\phi$ is non-decreasing and $\phi(0) = 0$, we have the following possibilities for $\operatorname{dom}\phi$:

$$\operatorname{dom}\phi = \begin{cases} (-\infty, \bar{x}), 0 < \bar{x} < +\infty, \\ (-\infty, \bar{x}], 0 \leq \bar{x} < +\infty, \\ \mathbb{R}. \end{cases}$$

Let us rule out the first possibility. Suppose by contradiction that $\operatorname{dom}\phi = (-\infty, \bar{x})$, where $0 < \bar{x} < +\infty$. Take a sequence of positive numbers $\{x_k : k \in \mathbb{N}\} \subseteq \operatorname{dom}\phi$ such that $\lim_{k\in\mathbb{N}} x_k = \bar{x}$. Since $\phi$ is l.s.c., using equation (2.5.62) we obtain that:

$$+\infty = \phi(\bar{x}) \leq \liminf_{k\in\mathbb{N}} \phi(x_k) \leq \liminf_{k\in\mathbb{N}} L(\phi)x_k = L(\phi)\bar{x} < +\infty \ (\rightarrow\leftarrow).$$

Finally, let us rule out the second possibility. Suppose by contradiction that $\operatorname{dom}\phi = (-\infty, \bar{x}]$, where $0 \leq \bar{x} < +\infty$. By Proposition 2.5.33, we have that:

$$\partial\phi(\bar{x}) = [-\phi'(\bar{x}; -1), \phi'(\bar{x}; 1)] \cap \mathbb{R}. \tag{2.5.63}$$

Observe that $\phi(x) = +\infty$, for all $x > \bar{x}$. So, it follows that $\phi'(\bar{x}; 1) = +\infty$. Moreover, by Proposition 2.5.35, we have that:

$$-\phi'(\bar{x}; -1) = \lim_{x\to\bar{x}-} \phi'(x; 1) \leq L(\phi). \tag{2.5.64}$$

The last inequality holds because $x < \bar{x}$ belongs to $\operatorname{int}(\operatorname{dom}\phi)$, so $\partial\phi(x) \neq \emptyset$ which implies that $\phi'(x; 1) = \sup \partial\phi(x) \leq \sup \mathcal{I}_\phi = L(\phi)$[31]. It follows that $\partial\phi(\bar{x}) \supseteq [L(\phi), +\infty)$, so:

$$+\infty > L(\phi) = \sup \mathcal{I}_\phi \geq \sup \partial\phi(\bar{x}) = +\infty \ (\rightarrow\leftarrow),$$

and the result is proved. □

The following example shows that we really have to be careful with the steps made in the proof of the previous proposition.

**Example 2.5.39.** *Consider the following convex function:*

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R} \cup \{+\infty\} \\ x &\mapsto \begin{cases} x, & \textit{for } x < 1, \\ 2, & \textit{for } x = 1, \\ +\infty, & \textit{for } x > 1. \end{cases} \end{aligned}$$

*Observe that $f$ satisfies all properties defining $\Phi$, excepting the lower semicontinuity. Moreover,*

$$\partial f(x) = \begin{cases} \{1\}, & \textit{for } x < 1, \\ \emptyset, & \textit{otherwise.} \end{cases} \tag{2.5.65}$$

[31] It is not true, in general, that $f'(x; 1) = \sup \partial f(x)$, for $x \in \operatorname{dom} f$. See Example 2.5.39.

*We conclude that* $I_f = \cup_{x \in \operatorname{dom} f} \partial f(x) = \{1\}$*. Although* $L(f) = 1$*, it is not true that* $f$ *is Lipschitz continuous, even on* $\operatorname{dom} f$*! Also, observe that* $f'(1;1) = +\infty \neq \sup \partial f(1) = -\infty$. □

The following theorem is a classical result from real analysis.

**Theorem 2.5.40.** *(fundamental theorem of calculus for convex functions) Let* $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ *be a proper convex function satisfying* $\operatorname{int} \operatorname{dom} f \neq \emptyset$*. We have that* $f$ *is almost everywhere differentiable with respect to (w.r.t.) the Lebesgue measure on* $\operatorname{int} \operatorname{dom} f$*. Moreover, the left and right derivatives of* $f$ *are (finite) monotonically non-decreasing functions on* $\operatorname{int} \operatorname{dom} f$ *and* $f'(x)$ *is well-defined at* $x \in \operatorname{int} \operatorname{dom} f$ *if and only if* $-f'(x;-1) = f'(x;1)$*. Finally, for every* $x, y \in \operatorname{int} \operatorname{dom} f$ *satisfying* $x < y$*, we have that:*

$$f(y) - f(x) = \int_x^y f'(s)ds = \int_x^y f'(s;1)ds = \int_x^y -f'(s;-1)ds. \tag{2.5.66}$$

*Proof.* See [60, Page 110, Theorem 14 and Corollary 15; Page 113, Proposition 17]. □

Next we show an optimality condition for proper convex functions. First of all, let us recall that for a proper convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ we have that $\bar{x} \in \operatorname{argmin}_{x \in \mathbb{R}^n} f(x)$ if and only if $0 \in \partial f(\bar{x})$. Indeed, this follows immediately from Definition 2.5.20.

**Proposition 2.5.41.** *Let* $f : \mathbb{R} \to \mathbb{R}$ *be a convex function. We have that*

$$\operatorname*{argmin}_{x \in \mathbb{R}} f(x) = \mathbb{R} \cap [\underline{x}, \overline{x}], \tag{2.5.67}$$

*where*

$$\underline{x} := \sup\{x \in \mathbb{R} : -f'(x;-1) < 0\}, \tag{2.5.68}$$
$$\overline{x} := \inf\{x \in \mathbb{R} : f'(x;1) > 0\}. \tag{2.5.69}$$

*Proof.* We begin by showing the inclusion $\operatorname{argmin}_{x \in \mathbb{R}} f(x) \subseteq \mathbb{R} \cap [\underline{x}, \overline{x}]$. Of course, if $\operatorname{argmin}_{x \in \mathbb{R}} f(x) = \emptyset$, then there is nothing to prove. So, let $\tilde{x} \in \operatorname{argmin}_{x \in \mathbb{R}} f(x)$ be given. We just have to show that $\underline{x} \leq \tilde{x} \leq \overline{x}$. By Corollary 2.5.33 and the finiteness of $f$, we conclude that $0 \in \partial f(\tilde{x}) = [-f'(\tilde{x};-1), f'(\tilde{x};1)]$, i.e. $-f'(\tilde{x};-1) \leq 0 \leq f'(\tilde{x};1)$. If $x < \tilde{x}$, then $f'(x;1) \leq -f'(\tilde{x};-1) \leq 0$. It follows that $\{x \in \mathbb{R} : f'(x;1) > 0\} \subseteq [\tilde{x}, +\infty)$, i.e. $\tilde{x} \leq \overline{x}$. Moreover, if $x > \tilde{x}$, then $-f'(x;-1) \geq f'(\tilde{x};1) \geq 0$. It follows that $\{x \in \mathbb{R} : -f'(x;-1) < 0\} \subseteq (-\infty, \tilde{x}]$, i.e. $\underline{x} \leq \tilde{x}$.

Now, let us show the following inclusion: $\mathbb{R} \cap [\underline{x}, \overline{x}] \subseteq \operatorname{argmin}_{x \in \mathbb{R}} f(x)$. If $\mathbb{R} \cap [\underline{x}, \overline{x}] = \emptyset$, then there is nothing to prove. Let $\tilde{x} \in \mathbb{R} \cap [\underline{x}, \overline{x}]$ be arbitrary. We claim that $-f'(\tilde{x};-1) \leq 0$ and $f'(\tilde{x};1) \geq 0$, i.e. $0 \in \partial f(\tilde{x})$. Suppose, by contradiction, that $-f'(\tilde{x};-1) > 0$. By Proposition 2.5.35, we have that

$$-f'(\tilde{x};-1) = \sup_{x<\tilde{x}} f'(x;1). \tag{2.5.70}$$

It follows that $f'(x;1) > 0$, for some $x < \tilde{x}$, i.e. $\overline{x} \leq x < \tilde{x}$ ($\rightarrow\leftarrow$). Similarly, suppose by contradiction that $f'(\tilde{x};1) < 0$. Again, by Proposition 2.5.35, we have that

$$f'(\tilde{x};1) = \inf_{x>\tilde{x}} -f'(x;-1). \tag{2.5.71}$$

It follows that $-f'(x;-1) < 0$, for some $x > \tilde{x}$, i.e. $\underline{x} \geq x > \tilde{x}$ ($\rightarrow\leftarrow$). The proposition is proved. $\square$

The following result is particularly useful to study the set of optimal solutions of the optimization problem associated with an OCE risk measure.

**Proposition 2.5.42.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a finite-valued convex function. If $Z$ is a random variable such that:*

$$g(t) := \mathbb{E}f(Z-t) \in \mathbb{R}, \tag{2.5.72}$$

*for all $t \in \mathbb{R}$, then $g$ is a convex function that satisfies:*

$$\partial g(t) = \left[-\mathbb{E}f'(Z-t;1), \mathbb{E}f'(Z-t;-1)\right], \tag{2.5.73}$$

*for all $t \in \mathbb{R}$.*

*Proof.* This result follows in particular from [73, Theorem 7.51]. $\square$

Observe that if $Z$ is an integrable random variable and $f$ is Lipschitz continuous, then condition (2.5.74) of Proposition 2.5.42 is satisfied.

**Proposition 2.5.43.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a $L$-Lipschitz continuous function. If $Z$ is an integrable random variable, then:*

$$g(t) := \mathbb{E}f(Z-t) \in \mathbb{R}, \tag{2.5.74}$$

*for all $t \in \mathbb{R}$.*

*Proof.* We have that:

$$|f(Z-t)| - |f(-t)| \leq |f(Z-t) - f(-t)| \leq L|(Z-t)-(-t)| \leq L|Z|. \tag{2.5.75}$$

It follows that:

$$|f(Z-t)| \leq L|Z| + |f(-t)|, \tag{2.5.76}$$

and so:

$$\mathbb{E}|f(Z-t)| \leq L\mathbb{E}|Z| + |f(-t)| < +\infty. \tag{2.5.77}$$

We conclude that $g(t) = \mathbb{E}f(Z-t) \in \mathbb{R}$, for all $t \in \mathbb{R}$. $\square$

The following result shows that a sort of reversed inequality holds for convex functions when we deal with an affine combination of elements that are not a convex combination.

**Proposition 2.5.44.** *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function. Given $x_1, x_2 \in \mathbb{R}^n$, we have that:*

$$f(\lambda x_1 + (1-\lambda)x_2) \geq \lambda f(x_1) + (1-\lambda) f(x_2), \tag{2.5.78}$$

*for $\lambda < 0$ or $\lambda > 1$.*

*Proof.* Suppose that $\lambda > 1$. Observe that $0 < 1/\lambda < 1$ and

$$x_1 = \frac{1}{\lambda}\left(\lambda x_1 + (1-\lambda)x_2\right) + \left(1 - \frac{1}{\lambda}\right) x_2, \tag{2.5.79}$$

so

$$f(x_1) \leq \frac{1}{\lambda} f\left(\lambda x_1 + (1-\lambda)x_2\right) + \left(1 - \frac{1}{\lambda}\right) f(x_2), \tag{2.5.80}$$

since $f$ is convex. Multiplying the previous equation by $\lambda$, we obtain, after some algebra, that equation (2.5.78) holds for $\lambda > 1$. For $\lambda < 0$, apply the same reasoning to $1 - \lambda > 1$, by writing $x_2$ as a convex combination of $x_1$ and $\lambda x_1 + (1-\lambda)x_2$. □

## 2.6 Set-valued analysis

In this section we make a short presentation of set-valued analysis, introducing the concept of continuity of a *multifunction*[32] between metric spaces. This section has two subsections. In the first one we present some results in parametric optimization. We study under which conditions on the problem data the optimal value function and the multifunction of optimal solutions are, respectively, continuous and outer semicontinuous. In the second subsection we discuss measurable multifunctions and present sufficient conditions for the measurability of optimal value functions. The results presented here are standard ones in this somewhat non-standard field of mathematics. For a more detailed discussion about these topics, the reader should consult [13, 59, 6].

Let $(X, d)$ be a metric space[33]. Let us consider two concepts of set convergence in metric spaces.

[32] Also known in the literature as a *set-valued mapping*, *point-to-set mapping* or *correspondence*.

[33] Usually we do not mention the metric $d$ explicitly and just say that $X$ is a metric space. Let us recall that a metric in $X$ is a function $d : X \times X \to \mathbb{R}_+$ satisfying the following properties: (i) $d(x, y) = 0 \Leftrightarrow x = y$, (ii) $d(x, y) = d(y, x)$, for all $x, y \in X$, and (c) for every $x, y, z \in X$, $d(x, z) \leq d(x, y) + d(y, z)$.

**Definition 2.6.1.** *Let $\{A_k : k \in \mathbb{N}\}$ be a sequence of subsets of $X$, i.e. $A_k \subseteq X$ for all $k \in \mathbb{N}$. Let us define, respectively, the interior and exterior limits*[34] $\{A_k : k \in \mathbb{N}\}$ *by:*

$$\operatorname*{limint}_{k\in\mathbb{N}} A_k := \left\{ x \in X : \limsup_{k\in\mathbb{N}} d(x, A_k) = 0 \right\}, \tag{2.6.1}$$

$$\operatorname*{limext}_{k\in\mathbb{N}} A_k := \left\{ x \in X : \liminf_{k\in\mathbb{N}} d(x, A_k) = 0 \right\}. \tag{2.6.2}$$

It is elementary to verify that $\operatorname*{limint}_{k\in\mathbb{N}} A_k \subseteq \operatorname*{limext}_{k\in\mathbb{N}} A_k$. When both limits coincide, that is:

$$A := \operatorname*{limext}_{k\in\mathbb{N}} A_k = \operatorname*{limint}_{k\in\mathbb{N}} A_k, \tag{2.6.3}$$

we say that the sequence $\{A_k : k \in \mathbb{N}\}$ converges to $A$. Now we consider the definition of a *multifunction.*

**Definition 2.6.2.** *(multifunction) Consider $X$ and $Y$ (nonempty) sets. A multifunction $S : X \rightrightarrows Y$ associates every input $x \in X$ with an output $S(x)$ that is a subset of $Y$, i.e. $S(x) \subseteq Y$. We define the domain of $S$ by:*

$$\operatorname{dom} S := \{x \in X : S(x) \neq \emptyset\}. \tag{2.6.4}$$

Now, we recall the concepts of inner-semicontinuity (I.S.C.), outer-semicontinuity (O.S.C.) and continuity of a multifunction.

**Definition 2.6.3.** *Let $S : X \rightrightarrows Y$ be a multifunction between (nonempty) metric spaces $X$ and $Y$. Let $x \in X$ be arbitrary.*

*(a) We say that $S$ is O.S.C. at $x$ if whenever a sequence $\{x_k : k \in \mathbb{N}\} \subseteq X$ converges to $x$, then*

$$\operatorname*{limext}_{k\in\mathbb{N}} S(x_k) \subseteq S(x).$$

*(b) We say that $S$ is I.S.C. at $x$ if whenever a sequence $\{x_k : k \in \mathbb{N}\} \subseteq X$ converges to $x$, then*

$$S(x) \subseteq \operatorname*{limint}_{k\in\mathbb{N}} S(x_k).$$

*(c) We say that $S$ is continuous at $x$ if $S$ is O.S.C. and I.S.C. at $x$.*

We also consider the restriction of a multifunction $S$ to an arbitrary (nonempty) subset $W$ of $X$:

$$\begin{aligned} S_{|W} : \quad W &\rightrightarrows Y \\ x &\mapsto S(x) \subseteq Y. \end{aligned} \tag{2.6.5}$$

Whenever it is said that $S$ restricted to $W$ is I.S.C. or O.S.C. at $x \in W$, we consider only sequences $\{x_k : k \in \mathbb{N}\}$ converging to $x$ that are on $W$.

[34]Some authors denote these limits as $\liminf_{k\in\mathbb{N}} A_k$ and $\limsup_{k\in\mathbb{N}} A_k$, respectively. We prefer to use another notation in this thesis in order to avoid any possible confusion with the identical well-established notation adopted in measure theory, i.e., $\liminf_{k\in\mathbb{N}} A_k = \cup_{j\in\mathbb{N}} \cap_{k\geq j} A_k$ and $\limsup_{k\in\mathbb{N}} A_k = \cap_{j\in\mathbb{N}} \cup_{k\geq j} A_k$.

### 2.6.1 Continuity of optimal value functions

In this section, unless stated otherwise, $X$ and $\Theta$ are nonempty metric spaces, $g : X \times \Theta \to \mathbb{R}$ is a function, and $C : \Theta \rightrightarrows X$ is a multifunction. Consider the optimal value function

$$h(\theta) := \inf_{x \in C(\theta)} g(x, \theta), \text{ for all } \theta \in \Theta, \tag{2.6.6}$$

and the multifunction $S : \Theta \rightrightarrows X$ of optimal solutions

$$S(\theta) := \operatorname*{argmin}_{x \in C(\theta)} g(x, \theta), \text{ for all } \theta \in \Theta. \tag{2.6.7}$$

We are interested in studying under which conditions in the problem data, function $h$ is continuous at $\theta_0 \in \Theta$. As a bonus we also obtain that under these same conditions the multifunction $S$ is O.S.C. at $\theta_0$. There are at least two possible ways to derive these kind of results. Here we present a version of Berge's Maximum Theorem (BMT) [6, Page 116]. For a different approach one should consult [73, Section 7.1.5] or [10, Proposition 4.4].

We apply the results of this section for showing that under appropriate regularity conditions w.p.1 the SAA problem in the multistage setting is such that

$$x_1 \in X_1 \mapsto \hat{\mathcal{Q}}_2(x_1) = \sum_{j=1}^{N_2} \hat{Q}_2(x_1, \xi^j) \tag{2.6.8}$$

is a continuous function (see Proposition 2.1.15). The continuity of $\hat{\mathcal{Q}}_2(\cdot) : X_1 \to \mathbb{R}$ was shown assuming that the multifunctions $X_{t+1}(\cdot, \xi_{t+1}) : \mathcal{X}_t \rightrightarrows \mathbb{R}^{n_{t+1}}$ are continuous, for all $\xi_{t+1} \in \operatorname{supp} \xi_{t+1}$, $t = 1, \ldots, T-1$ (see conditions (Mt.5), for $t = 1, \ldots, T-1$ in Section 2.1.2).

Proposition 2.6.4 is a version (see Remark 2.6.5) of BMT.

**Proposition 2.6.4.** *Take any $\theta_0 \in \Theta$. Suppose that the following conditions hold: (i) the function $g : X \times \Theta \to \mathbb{R}$ is continuous, (ii) the multifunction $C : \Theta \rightrightarrows X$ is compact-valued, (iii) there exists a neighborhood $V$ of $\theta_0 \in \operatorname{dom}(C)$ such that*

$$C(V) = \bigcup_{\theta \in V} C(\theta) \tag{2.6.9}$$

*is a compact metric space, and (iv) $C$ is continuous at $\theta_0$. Then $h$ is continuous at $\theta_0$ and $S$ is O.S.C. at $\theta_0$.*

*Proof.* First, let us show that $\theta_0 \in \operatorname{int}(\operatorname{dom}(C))$. Suppose by contradiction that $\theta_0 \notin \operatorname{int}(\operatorname{dom}(C))$. Thus, there exists a sequence $\{\theta_k : k \in \mathbb{N}\} \subseteq \Theta \setminus \operatorname{dom}(C)$ such that $\lim_k \theta_k = \theta_0$. Note that $C(\theta_k) = \emptyset$, for all $k \in \mathbb{N}$, so

$$\emptyset \neq C(\theta_0) \subseteq \operatorname*{lim\,int}_k C(\theta_k) = \emptyset \ (\rightarrow\leftarrow).$$

Now, let $\{\theta_k : k \in \mathbb{N}\}$ be any sequence in $\Theta$ converging to $\theta_0$. For proving that $h(\theta_0) = \lim_k h(\theta_k)$, it is sufficient to show that for any subsequence $\{\theta_k : k \in N'\}$, where $N' \subseteq \mathbb{N}$, there exists a subsubsequence $\{\theta_k : k \in N''\}$, $N'' \subseteq N'$, such that $h(\theta_0) = \lim_{k\in N''} h(\theta_k)$. Observe that there exists $K \in \mathbb{N}$ such that $\theta_k \in V \cap \mathrm{dom}(C)$, for all $k \geq K$. Since $C$ is compact-valued and $g$ is continuous, we can consider a selection $x_k \in S(\theta_k) \subseteq C(\theta_k)$, for $k \geq K$ and $k \in N'$. Since $C(V)$ is compact and $x_k \in C(V)$, for all $k \geq K$, there exists a subsequence $\{x_k : k \in N''\}$, where $N'' \subseteq N'$, such that $x_k \longrightarrow \bar{x} \in C(V)$, as $k \in K''$ goes to $+\infty$. Note that $\bar{x} \in \mathrm{limext}_k\, C(\theta_k) \subseteq C(\theta_0)$, where the inclusion holds since $C$ is O.S.C. at $\theta_0$. Moreover, let $x \in C(\theta_0)$ be arbitrary. Since $C$ is I.S.C. at $\theta_0$, there exists $y_k \in C(\theta_k)$ (for $k \geq K$) such that $y_k \longrightarrow x$, as $k \to +\infty$. Summing up, we obtain that

$$g(x_k, \theta_k) \leq g(y_k, \theta_k), \text{ for all } k \in N'' \text{ and } k \geq K.$$

Letting $k \to +\infty$ above and using the (jointly) continuity of $g$, we obtain that $g(\bar{x}, \theta_0) \leq g(x, \theta_0)$, for all $x \in C(\theta_0)$. This shows that $\bar{x} \in S(\theta_0)$ and that $h(\theta_0) = \lim_{k\in N''} h(\theta_k)$, proving the continuity of $h$ at $\theta_0$.

Finally, let us show that $\mathrm{limext}_k\, S(\theta_k) \subseteq S(\theta_0)$. Take any $\hat{x} \in \mathrm{limext}_k\, S(\theta_k)$. It follows that there exists a subsequence $\{x_{k_j} : j \in \mathbb{N}\}$ such that $x_{k_j} \longrightarrow \hat{x}$, as $j \to +\infty$, where $x_{k_j} \in S(\theta_{k_j})$, for all $k \in \mathbb{N}$. Take any $x \in C(\theta_0)$. Using again the fact that $C$ is I.S.C. at $\theta_0$, there exists a sequence $\{y_k : k \geq K\}$ converging to $x$ such that $y_k \in C(\theta_k)$, for all $k \geq K$. Using again the continuity of $g$, we obtain that $g(\hat{x}, \theta_0) \leq g(x, \theta_0)$. It follows that $\hat{x} \in S(\theta_0)$ and the result is proved. □

**Remark 2.6.5.** *Let us make two remarks about the differences between our presentation of BMT and the original theorem. First note that here we present the result considering a parametric minimization problem, instead of a parametric maximization problem. This difference is irrelevant in the sense that trivial modifications of the proofs presented here and in the original versions of BMT work, respectively, for parametric maximization problems and for parametric minimization problems. A more relevant difference between the versions is that in [6] the author considers a different notion of continuity of a multifunction that the one considered here. In [6, Chapter VI, §1] the author defines the concepts of lower semicontinuous (l.s.c.) and upper semicontinuous (u.s.c.) multifunctions at a given point of its domain. He states that a multifunction is continuous at a given point if it is l.s.c. and u.s.c. at this point. It is possible to show that the concept of a l.s.c. multifunction considered in [6] is equivalent to the concept of a I.S.C. multifunction considered here. However, the concept of a u.s.c. multifunction considered in [6] is different from the concept of a O.S.C. multifunction considered here. In [13] the authors considered, besides the notion of continuity of a multifunction presented here, the concept of Kuratowski-continuity. This concept coincides with the one presented in [6].* □

In the next proposition we consider metrics on the subsets $X$ and $Y$ of the Euclidean spaces $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, that are given by any norm of $\mathbb{R}^n$ and $\mathbb{R}^m$.

**Proposition 2.6.6.** *Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be (nonempty) compact sets. If $C : X \rightrightarrows Y$ is a continuous multifunction, then $C$ is compact-valued, $C(X) := \cup_{x \in X} C(x) \subseteq \mathbb{R}^m$ is compact and* $\mathrm{dom}(C)$ *is open in $X$ and compact.*

*Proof.* If $\mathrm{dom}(C) = \emptyset$, the result holds trivially. It is well-known (see [13, Theorem 2.5.4]) that $C : X \rightrightarrows Y$ is O.S.C. at $X$ if and only if $\mathrm{gph}(C) \subseteq X \times Y$ is closed in $X \times Y$. Any closed subset of a compact set is compact. Note that $X \times Y$ is compact, so $\mathrm{gph}(C)$ is also compact. Denote by $\pi_x(x, y) := x$ and $\pi_y(x, y) := y$ the projections in the $x$-variable and $y$-variable, respectively. Since these functions are continuous, it follows that $\mathrm{dom}(C) = \pi_x(\mathrm{gph}(C))$ and $C(X) = \pi_y(\mathrm{gph}(C))$ are compact sets. In particular, $C(x) \subseteq \mathbb{R}^m$ is compact, for all $x \in X$. Let us show that $\mathrm{dom}(C)$ is open in $X$. Suppose by contradiction that $\mathrm{dom}(C)$ is not open in $X$ and take any $x \in \mathrm{dom}(C) \setminus \mathrm{int}_X(\mathrm{dom}(C))$. So, there exists a sequence $\{x^k : k \in \mathbb{N}\}$ in $X \setminus \mathrm{dom}\, C$ such that $x^k \longrightarrow x$, as $k \to +\infty$. Since $C$ is I.S.C. at $x$, we obtain that $C(x) \subseteq \mathrm{limint}_k\, C(x^k) = \emptyset$, which contradicts the fact that $x \in \mathrm{dom}(C)$. $\square$

In many applications the multifunction $C : \Theta \rightrightarrows \mathbb{R}^n$ is of the form

$$C(\theta) := \{x \in \mathbb{R}^n : G(x, \theta) \in K\}, \tag{2.6.10}$$

where $G : \mathbb{R}^n \times \Theta \to \mathbb{R}^m$ is a function, $K \subseteq \mathbb{R}^m$ is a pointed closed convex cone, and $\Theta$ is a metric space. In Proposition 2.6.7 we present some sufficient conditions that guarantee the outer semicontinuity and the inner semicontinuity of $C(\cdot)$ at $\theta_0 \in \Theta$.

**Proposition 2.6.7.** *Take any nonempty metric space $\Theta$. Let $C : \Theta \rightrightarrows \mathbb{R}^n$ be a multifunction of the form (2.6.10) where $K \subseteq \mathbb{R}^m$ is a closed convex cone (see Definition 2.5.5) and $G : \mathbb{R}^n \times \Theta \to \mathbb{R}^m$ is a function. The following assertions hold: (a) if $G$ is continuous, then $C$ is O.S.C. in $\Theta$, (b) suppose, additionally, that $G(\cdot, \theta)$ is convex with respect to $-K$ (see Definition 2.5.18) and that the Slater constraint qualification is satisfied at $\theta_0 \in \Theta$, i.e. there exists $x_0 \in \mathbb{R}^n$ such that $G(x_0, \theta_0) \in \mathrm{int}\, K$. Then, $C(\cdot)$ is I.S.C. at $\theta_0$.*

*Proof.* For proving item (a) it is sufficient (and necessary) to show that $\mathrm{gph}(C)$ is closed in $\Theta \times \mathbb{R}^n$. Take any sequence $(\theta_j, x_j) \in \mathrm{gph}(C)$ such that $(\theta_j, x_j) \to (\bar{\theta}, \bar{x}) \in \Theta \times \mathbb{R}^n$. We just need to show that $(\bar{\theta}, \bar{x}) \in \mathrm{gph}\, C$, i.e. $\bar{x} \in C(\bar{\theta})$. Note that $G(x_j, \theta_j) \in K$, for every $j \in \mathbb{R}^n$. Moreover, since $G$ is continuous and $(\theta_j, x_j) \to (\bar{\theta}, \bar{x})$, it follows that $G(\bar{x}, \bar{\theta}) = \lim_j G(x_j, \theta_j) \in \overline{K} = K$, i.e. $\bar{x} \in C(\bar{\theta})$. This proves item (a).

Now we prove item (b). Take any sequence $\theta_j \to \theta_0$. For showing that $C(\cdot)$ is I.S.C. at $\theta_0$, we just have to prove that $C(\theta_0) \subseteq \operatorname{limint}_j C(\theta_j)$. Take any $\hat{x} \in C(\theta_0)$. We consider two steps.

Step 1: Suppose that $G(\hat{x}, \theta_0) \in \operatorname{int} K$. Note that $G(\hat{x}, \theta_j) \to_j G(\hat{x}, \theta_0) \in \operatorname{int} K$, since $G$ is continuous. Therefore, $G(\hat{x}, \theta_j) \in K$ for sufficiently large $j \in \mathbb{N}$, that is $\hat{x} \in C(\theta_j)$ for sufficiently large $j$. It follows that $\hat{x} \in \operatorname{limint}_j C(\theta_j)$.

Step 2: Take any $0 \leq \lambda < 1$. By hypothesis, there exists $x_0 \in \mathbb{R}^n$ such that $G(x_0, \theta_0) \in \operatorname{int} K$. Note that $\lambda G(\hat{x}, \theta_0) + (1-\lambda) G(x_0, \theta_0) \in \operatorname{int} K$, since $1 - \lambda > 0$ and $G(\hat{x}, \theta_0) \in K$. Since $G(\cdot, \theta_0)$ is convex with respect to $-K$, we have that

$$\lambda G(\hat{x}, \theta_0) + (1-\lambda) G(x_0, \theta_0) \succeq_{-K} G(\lambda \hat{x} + (1-\lambda) x_0, \theta_0), \tag{2.6.11}$$

that is:

$$G(\lambda \hat{x} + (1-\lambda) x_0, \theta_0) - [\lambda G(\hat{x}, \theta_0) + (1-\lambda) G(x_0, \theta_0)] \in K. \tag{2.6.12}$$

Since $K$ is a convex cone and $\lambda G(\hat{x}, \theta_0) + (1-\lambda) G(x_0, \theta_0) \in \operatorname{int} K$, it follows from Lemma 2.5.9 that

$$\begin{aligned} G(\lambda \hat{x} + (1-\lambda) x_0, \theta_0) = (G(\lambda \hat{x} + (1-\lambda) x_0, \theta_0) - [\lambda G(\hat{x}, \theta_0) + (1-\lambda) G(x_0, \theta_0)]) \\ + [\lambda G(\hat{x}, \theta_0) + (1-\lambda) G(x_0, \theta_0)] \in \operatorname{int} K. \end{aligned}$$

By Step 1, we conclude that $\lambda \hat{x} + (1-\lambda) x_0 \in \operatorname{limint}_j C(\theta_j)$, for $0 \leq \lambda < 1$. Letting $\lambda \to 1$, we obtain that $\hat{x} \in \overline{\operatorname{limint}_j C(\theta_j)} = \operatorname{limint}_j C(\theta_j)$, since $\operatorname{limint}_j C(\theta_j)$ is a closed set (see [13, Remark 2.5.2]). This completes the proof of item (b). □

### 2.6.2 Measurability of multifunctions

In this section we follow closely reference [59, Chapter 14]. Another reference on this topic is [73, Section 7.2.3]. The main objective of this section is to introduce the basic theory of measurable multifunctions that guarantees that optimal value functions of the form

$$q(\omega) := \inf_{x \in \mathbb{R}^n} f(x, \omega) \tag{2.6.13}$$

are $\mathcal{F}$-measurable, where $(\Omega, \mathcal{F})$ is a measurable space and $f : \mathbb{R}^n \times \Omega \to \overline{\mathbb{R}}$ is a extended real-valued function. Although it does not appear any multifunction in the formulation above, the sufficient conditions on $f$ that guarantee the measurability of $q(\cdot)$ will involve the notion of measurable multifunctions. This topic is particularly useful when we deal with technical measurability questions of two-stage and multistage stochastic programming problems. Let us begin by recalling the definition of a measurable multifunction. In the sequel, $(\Omega, \mathcal{F})$ will always be a measurable space.

**Definition 2.6.8.** *(measurable multifunctions) A multifunction $S : \Omega \rightrightarrows \mathbb{R}^n$ is measurable if for every open set $G \subseteq \mathbb{R}^n$ its inverse image by $S$*

$$S^{-1}(G) := \{\omega \in \Omega : S(\omega) \cap G \neq \emptyset\} \tag{2.6.14}$$

*belongs to $\mathcal{F}$.*

Proposition 2.6.9 states equivalent conditions for the measurability of a multifunction $S : \Omega \rightrightarrows \mathbb{R}^n$. Here we state just some equivalent statements that will be useful in our presentation. See [59, Theorem 14.3] for other equivalent conditions.

**Proposition 2.6.9.** *Let $S : \Omega \rightrightarrows \mathbb{R}^n$ be a multifunction. Consider the following statements:*

*(a) $S^{-1}(G) \in \mathcal{F}$, for all open sets $G \subseteq \mathbb{R}^n$.*

*(b) $S^{-1}(F) \in \mathcal{F}$, for all closed sets $F \subseteq \mathbb{R}^n$.*

*(c) the function $\omega \in \Omega \to d\,(x, S(\omega))$ is $\mathcal{F}$-measurable for each $x \in \mathbb{R}^n$.*

*We have that (a) $\Leftrightarrow$ (b) $\Rightarrow$ (c). Moreover, when $S$ is closed-valued, we also have that (c) $\Rightarrow$ (a).*

*Proof.* See [59, Theorem 14.3]. □

**Remark 2.6.10.** *Let $S : \Omega \rightrightarrows \mathbb{R}^n$ be a measurable multifunction. Differently from measurable functions, it is not true, in general, that $S^{-1}(B) \in \mathcal{F}$, for every $B \in \mathbb{B}(\mathbb{R}^n)$. See [59, Theorem 14.8] for more details about this topic.* □

The following result will be useful.

**Proposition 2.6.11.** *Consider $J$ a countable index set and let $S_j : \Omega \rightrightarrows \mathbb{R}^n$ be measurable, for each $j \in J$. The following assertions hold true:*

*(a) $\omega \in \Omega \mapsto \bigcup_{j\in J} S_j(\omega)$ is measurable,*

*(b) if each $S_j$ is closed-valued, then $\omega \in \Omega \mapsto \bigcap_{j\in J} S_j(\omega)$ is measurable.*

*(c) if $J$ is finite, say $J = \{1, \dots, m\}$, then $\omega \in \Omega \mapsto S_1(\omega) \times \cdots \times S_m(\omega)$ is measurable.*

*Proof.* See [59, Proposition 14.11]. □

Given a function $f : \mathbb{R}^n \times \Omega \to \overline{R}$, we consider its epigraph multifunction

$$E_f : \omega \in \Omega \mapsto E_f(\omega) := \operatorname{epi} f(\cdot, \omega) \subseteq \mathbb{R}^{n+1}. \tag{2.6.15}$$

Note that for any $\omega \in \Omega$, $f(\cdot, \omega)$ is the function that associates every input $x \in \mathbb{R}^n$ with the output $f(x, \omega) \in \overline{\mathbb{R}}$. So, $\operatorname{epi} f(\cdot, \omega) \subseteq \mathbb{R}^{n+1}$, for every $\omega \in \Omega$.

**Definition 2.6.12.** *(normal integrands or random l.s.c. functions) We say that $f : \mathbb{R}^n \times \Omega \to \overline{\mathbb{R}}$ is a normal integrand (or random l.s.c. function) when $E_f$ is closed-valued and measurable.*

**Remark 2.6.13.** *If $f$ is a normal integrand, then $\operatorname{epi} f(\cdot, \omega)$ is a closed set, i.e. $f(\cdot, \omega)$ is a l.s.c. function. It is also true that $f(x, \cdot)$ is $\mathcal{F}$-measurable, for every $x \in \mathbb{R}^n$. Let us show that. Take any $x \in \mathbb{R}^n$. We just have to verify that*

$$f^{-1}(x, F) = \{\omega \in \Omega : f(x, \omega) \in F\} \in \mathcal{F}, \tag{2.6.16}$$

*for every closed set $F \subseteq \mathbb{R}$. Note that*

$$f^{-1}(x, F) = E_f^{-1}(\{x\} \times F) \tag{2.6.17}$$

*that belongs to $\mathcal{F}$ since $E_f$ is a measurable multifunction and $\{x\} \times F$ is closed (see Proposition 2.6.9).*

*For closing this remark, let us point out that it is not true, in general, that a function $f : \mathbb{R}^n \times \Omega \to \overline{\mathbb{R}}$ satisfying conditions (i) and (ii) below is a normal integrand, where:*

*(i) $f(x, \cdot)$ is measurable, for all $x \in \mathbb{R}^n$,*

*(ii) $f(\cdot, \omega)$ is l.s.c., for every $\omega \in \Omega$.*

*For a counterexample see the proof of [59, Proposition 14.28].* □

Before proceeding we present the definition of Carathéodory functions. This is an important subclass of normal integrands (see Proposition 2.6.15).

**Definition 2.6.14.** *(Carathéodory functions) We say that $f : \mathbb{R}^n \times \Omega \to \mathbb{R}$ is a Carathéodory function when $f(x, \cdot)$ is $\mathcal{F}$-measurable for every $x \in \mathbb{R}^n$ and $f(\cdot, \omega)$ is continuous for every $\omega \in \Omega$.*

**Proposition 2.6.15.** *If $f : \mathbb{R}^n \times \Omega \to \mathbb{R}$ is a Carathéodory function, then $f$ is a normal integrand.*

*Proof.* Define the function $F : (\mathbb{R}^n \times \mathbb{R}) \times \Omega \to \mathbb{R}$ as

$$F(x, \alpha, \omega) := f(x, \omega) - \alpha. \tag{2.6.18}$$

Since $f$ is a Carathéodory function, the following assertions hold: (a) $(x, \alpha) \mapsto F(x, \alpha, \omega) = f(x, \omega) - \alpha$ is continuous, for every $\omega \in \Omega$, and (b) $\omega \in \Omega \mapsto F(x, \alpha, \omega) = f(x, \omega) - \alpha$ is $\mathcal{F}$-measurable, for every $(x, \alpha) \in \mathbb{R}^n \times \mathbb{R}$. Saying equivalently, $F$ is also a Carathéodory function. Now, note that

$$E_f(\omega) = F^{-1}(\cdot, \cdot, \omega)((-\infty, 0]), \tag{2.6.19}$$

since $(x, \alpha) \in F^{-1}(\cdot, \cdot, \omega)((-\infty, 0]) \Leftrightarrow F(x, \alpha, \omega) = f(x, \omega) - \alpha \leq 0 \Leftrightarrow f(x, \omega) \leq \alpha \Leftrightarrow (x, \alpha) \in \operatorname{epi} f(\cdot, \omega) \Leftrightarrow (x, \alpha) \in E_f(\omega)$. $E_f(\omega)$ is a closed set, since $F(\cdot, \cdot, \omega)$ is continuous and $(-\infty, 0]$ is closed. This proves that $E_f : \Omega \rightrightarrows \mathbb{R}^n \times \mathbb{R}$ is closed-valued.

Now, let us show that $E_f$ is a measurable multifunction. Take any open $G \subseteq \mathbb{R}^n \times \mathbb{R}$. We just need to prove that $E_f^{-1}(G) \in \mathcal{F}$. If $G = \emptyset$, then there is nothing to be done: $E_f^{-1}(\emptyset) = \emptyset \in \mathcal{F}$. Otherwise, let $D = \{(x_k, t_k) : k \in \mathbb{N}\}$ be a countable dense subset of $G$. We claim that

$$\begin{aligned} E_f^{-1}(G) &= E_f^{-1}(D) && (2.6.20)\\ &= \bigcup_{k\in\mathbb{N}} E_f^{-1}(\{(x_k, t_k)\}) && (2.6.21)\\ &= \bigcup_{k\in\mathbb{N}} f^{-1}(x_k, \cdot)\left((-\infty, t_k]\right) \in \mathcal{F}. && (2.6.22)\end{aligned}$$

It is elementary to verify that the second and third equalities hold. The countable union in the left-side of the third equality is a measurable set, since

$$f^{-1}(x_k, \cdot)\left((-\infty, t_k]\right) \in \mathcal{F},$$

for every $k \in \mathbb{N}$. Indeed, this follows from the facts that $(-\infty, t_k]$ is closed and $f(x_k, \cdot)$ is $\mathcal{F}$-measurable. Now, we show that the first equality holds. Since $G \supseteq D$, it follows that $E_f^{-1}(G) \supseteq E_f^{-1}(D)$. So, we just need to prove the converse inclusion. Take any $\omega \in E_f^{-1}(G)$. It follows that there exists $(x, t) \in E_f(\omega) \cap G$. Since $G$ is open, there exists $\epsilon > 0$ such that $B((x, t), \epsilon) \subseteq G$. Moreover, by the continuity of $f(\cdot, \omega)$, there exists $\delta > 0$, that we can take less than or equal to $\epsilon/2$, such that $f(x', \omega) < t + \epsilon/4$, for every $x' \in B(x, \delta)$. Finally, note that $B(x, \delta) \times (t + \epsilon/4, t + \epsilon/2) \subseteq B((x, t), \epsilon) \subseteq G$ is open. Therefore, $(x_k, t_k) \in B(x, \delta) \times (t + \epsilon/4, t + \epsilon/2)$, for some $k \in \mathbb{N}$. It follows that

$$f(x_k, \omega) < t + \epsilon/4 < t_k, \tag{2.6.23}$$

i.e. $(x_k, t_k) \in \operatorname{epi} f(\cdot, \omega) = E_f(\omega)$, in particular, $\omega \in E_f^{-1}(D)$. This completes the proof of the proposition. □

Now we present the key result of this subsection.

**Proposition 2.6.16.** *Consider a normal integrand $f : \mathbb{R}^n \times \Omega \to \overline{\mathbb{R}}$ and define:*

$$\begin{aligned} q(\omega) &:= \inf_{x\in\mathbb{R}^n} f(x, \omega); && (2.6.24)\\ S(\omega) &:= \operatorname*{argmin}_{x\in\mathbb{R}^n} f(x, \omega). && (2.6.25)\end{aligned}$$

*Then the optimal value function $q : \Omega \to \overline{\mathbb{R}}$ is $\mathcal{F}$-measurable and the multifunction of optimal solutions $S : \Omega \rightrightarrows \mathbb{R}^n$ is closed-valued and $\mathcal{F}$-measurable.*

*Proof.* See [59, Theorem 14.37]. □

**Corollary 2.6.17.** *Consider a normal integrand $f : \mathbb{R}^n \times \Omega \to \overline{\mathbb{R}}$ and a closed-valued measurable multifunction $X : \Omega \rightrightarrows \mathbb{R}^n$. Define:*

$$\begin{aligned} q(\omega) &:= \inf_{x \in X(\omega)} f(x,\omega); && (2.6.26) \\ S(\omega) &:= \operatorname*{argmin}_{x \in X(\omega)} f(x,\omega). && (2.6.27) \end{aligned}$$

*Then the optimal value function $q : \Omega \to \overline{\mathbb{R}}$ is $\mathcal{F}$-measurable and the multifunction of optimal solutions $S : \Omega \rightrightarrows \mathbb{R}^n$ is closed-valued and $\mathcal{F}$-measurable.*

*Proof.* Let us define the following function

$$\tilde{f}(x,\omega) := \begin{cases} f(x,\omega) & \text{if } x \in X(\omega), \\ +\infty & \text{otherwise.} \end{cases} \tag{2.6.28}$$

Note that $q(\omega) = \inf_{x\in\mathbb{R}^n} \tilde{f}(x,\omega)$ and $S(\omega) = X(\omega) \cap \operatorname{argmin}_{x\in\mathbb{R}^n} \tilde{f}(x,\omega)$. We claim that the result follows from Proposition 2.6.16 if we show that $\tilde{f}$ is a normal integrand. Indeed, assume that $\tilde{f}$ is a normal integrand. Proposition 2.6.16 implies that $q$ is measurable and

$$\omega \in \Omega \mapsto \operatorname*{argmin}_{x\in\mathbb{R}^n} \tilde{f}(x,\omega) \tag{2.6.29}$$

is closed-valued and measurable. Since $X$ is also closed-valued and measurable, it follows from Proposition 2.6.11 that

$$\omega \in \Omega \mapsto X(\omega) \cap \operatorname*{argmin}_{x\in\mathbb{R}^n} \tilde{f}(x,\omega) \tag{2.6.30}$$

is measurable and closed-valued.

Now, let us show that $\tilde{f}$ is a normal integrand. Note that

$$E_{\tilde{f}}(\omega) = E_f(\omega) \cap (X(\omega) \times \mathbb{R})\,. \tag{2.6.31}$$

It follows from Proposition 2.6.11 that $\omega \in \Omega \mapsto X(\omega) \times \mathbb{R}$ is closed-valued and measurable. Since $E_f$ is closed-valued and measurable, we conclude that $E_{\tilde{f}}$ is also closed-valued and measurable, i.e. $\tilde{f}$ is a normal integrand. □

## 2.7 Risk measures

In this section we present the notion of risk measures. As we have discussed previously, in some contexts the use of the expected value operator in order to summarize a random cost into a real number may not be a good criterion to be used in a stochastic optimization problem. That is particularly the case when lower and upper deviations of the random cost from its expected value do not perfectly offset in terms of the decision maker preferences. In a risk averse situation the relative losses of upper deviations of the random costs from its expected value are greater than the relative gains obtained from the lower deviations. In the sequel we consider many properties that a risk measure could present, like: risk aversity, convexity, coherence, law invariance, etc. We begin our presentation considering static risk measures, i.e. risk measures taking values on the real numbers. From static risk measures, it is straightforward to consider risk averse formulations of stochastic programming problems. Subsequently we consider conditional risk measures that are used to formulate risk averse multistage stochastic programming problems.

Unless stated otherwise, in the remainder of this section $(\Omega, \mathcal{F}, \mathbb{P})$ is a given probability space and $\mathcal{Z} = L_p(\Omega, \mathcal{F}, \mathbb{P})$ is the space of random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$ having finite $p$-th order moment, where $p \in [1, \infty)$. As usual we identify two random variables $Z$ and $W$ in $\mathcal{Z}$ that agree w.p.1. With some abuse of notation, we also write $Z \leq W$ to denote that $Z(\omega) \leq W(\omega)$ w.p.1.

**Definition 2.7.1.** *(risk measures) We say that $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is a risk measure on $\mathcal{Z}$ if it is a proper function, i.e.* $\operatorname{dom} \rho = \{Z \in \mathcal{Z} : \rho(Z) < +\infty\} \neq \emptyset$ *and $\rho(Z) > -\infty$, for all $Z \in \mathcal{Z}$.*

Since we are identifying two random variables $Z, W \in \mathcal{Z}$ that agree w.p.1, i.e., such that $\mathbb{P}[Z = W] = 1$, a risk measure $\rho$ defined on $\mathcal{Z}$ must relate the same value $\rho(Z) = \rho(W)$ for such random variables.

**Definition 2.7.2.** *(risk averse risk measures) We say that a risk measure $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is* risk averse *if it satisfies the following properties:*

*(a) $\rho(Z) \geq \mathbb{E}Z$, for all $Z \in \mathcal{Z}$.*

*(b) $\rho(a) = a$, for all $a \in \mathbb{R}$.*

**Remark 2.7.3.** *With some abuse of notation, we identify any real number $a \in \mathbb{R}$ with the constant random variable $a \cdot \mathbb{1}_\Omega(\cdot) \in \mathcal{Z}$. In that sense, $a \in \mathcal{Z}$ and $\rho(a)$ is well-defined, for every $a \in \mathbb{R}$.* □

Risk measures can be used by a decision maker to rank random outcomes. Take $Z$ and $W$ in $\mathcal{Z}$ and consider a risk measure $\rho$ defined in $\mathcal{Z}$. A decision maker that uses the risk measure $\rho$ as his choice criterion will choose $Z$ over $W$, whenever $\rho(Z) \leq \rho(W)$. By Definition 2.7.2 note that a risk averse decision maker prefers to incur in the certain cost $\mathbb{E}Z$ instead of incurring in the random cost $Z$, since

$$\rho(Z) \geq \mathbb{E}Z = \rho(\mathbb{E}Z), \tag{2.7.1}$$

for all $Z \in \mathcal{Z}$.

Definition 2.7.1 is too broad in the sense that it does not impose any consistency constraints on $\rho$. As we have pointed out previously, the elements of $\mathcal{Z}$ are random losses. One wants to compare distinct elements $Z, W \in \mathcal{Z}$ using a risk measure $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$. Of course, the greater a random loss $Z \in \mathcal{Z}$ is, the less desirable it must be. A reasonable risk measure $\rho$ must be able to meet this basic requirement condition. Let us consider the following set of axioms to be satisfied by a risk measure $\rho$:

(R1) Monotonicity: If $Z, W \in \mathcal{Z}$ and $Z \leq W$, then $\rho(Z) \leq \rho(W)$.

(R2) Translation equivariance: If $a \in \mathbb{R}$ and $Z \in \mathcal{Z}$, then $\rho(Z + a) = \rho(Z) + a$.

(R3) Convexity: For all $Z, W \in \mathcal{Z}$ and $\lambda \in [0, 1]$,

$$\rho(\lambda Z + (1 - \lambda)W) \leq \lambda\rho(Z) + (1 - \lambda)\rho(W).$$

(R4) Positive homogeneity: if $\lambda > 0$ and $Z \in \mathcal{Z}$, then $\rho(\lambda Z) = \lambda\rho(Z)$.

We say that a risk measure $\rho$ satisfying conditions (R1)-(R3) is a *convex* risk measure. If $\rho$ satisfies, additionally, condition (R4), we say that it is a *coherent* risk measure. The notion of a *coherent* risk measure was introduced in the seminal paper [3]. In [3] the authors stated that a risk measure $\rho$ is coherent if it satisfies conditions (R1), (R2), (R4) and

(R3') Subadditivity: For all $Z, W \in \mathcal{Z}$,

$$\rho(Z + W) \leq \rho(Z) + \rho(W).$$

It is elementary to show that the set of conditions (R1)-(R2)-(R3)-(R4) and (R1)-(R2)-(R3')-(R4) are equivalent. In [25] the authors dropped the last axiom (R4) from the list and substituted the axiom (R3') by (R3) to define the class of convex risk measures.

Now we argue that the axioms (R1)-(R3) are in fact quite natural conditions to be imposed to a risk measure $\rho$. Condition (R1) is obvious: if the random cost $Z$ is always less than or equal to $W$ (or even just w.p.1), then its perceived risk $\rho(Z)$ must be less than or equal to $\rho(W)$. Condition (R3) is related to portfolio diversification. It means that the risk associated with diversified positions must be at most equal to the weighted mean of the risks associated with the individual positions. Condition (R2) is natural to assume when we first consider the definition of acceptance sets $\mathcal{A} \subseteq \mathcal{Z}$ and then consider the risk measure $\rho_{\mathcal{A}}(\cdot)$ (see [25, Page 430]) constructed from $\mathcal{A}$ (see also the set of axioms (2.1)-(2.4) in [3] that must be satisfied by an acceptance set $\mathcal{A}$). Assume also that the risk measure $\rho$ satisfies the normalized condition $\rho(0) = 0$. In this case it follows that $\rho(a) = \rho(0 + a) = \rho(0) + a = a$, for all $a \in \mathbb{R}$. The quantity $\rho(Z)$ can be interpreted as the minimal amount of cash to be given to the decision maker in order for him to accept taking the random cost $Z$. As pointed out in [3], the axiom (2.4), that is related to the acceptance set $\mathcal{A}$, is a less natural condition to require. This axiom is intimately related with condition (R4) that was dropped in [25] when the authors considered the definition of convex risk measures. Indeed, in [25, Page 2] the authors argued that the risk associated with a given financial position might increase in a nonlinear way with the size of the position. In fact, an additional liquidity risk may arise if a position is multiplied by a large factor $\lambda > 0$.

As we have pointed before, in this thesis we derive sample complexity results for risk averse stochastic programming problems with OCE risk measures. It is worth mentioning that although the class of OCE risk measures is a subclass of convex risk measures (see Proposition 3.1.5), it is not true that every OCE risk measure is a coherent risk measure. However, by the discussion made in the last paragraph, this fact does not weaken the importance of this class of risk measure.

Now, let us consider some examples of risk measures (see [73, Section 6.3.2] for some other examples).

**Example 2.7.4.** *Let be given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Consider the following examples of risk measures:*

*(i) (Expected Value)* $\mathbb{E} : L_1(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$

$$\mathbb{E}Z = \int_\Omega Z(\omega) d\mathbb{P}(\omega). \tag{2.7.2}$$

*This is the most basic risk measure and it is used as a building block of many other important risk measures. It is elementary to verify that $\mathbb{E}(\cdot)$ is a coherent risk measure.*

*(ii) (Value-at-Risk) Take any $\alpha \in (0,1)$. The Value-at-Risk with significance level $\alpha$ of a random variable $Z$ is the left-side $(1-\alpha)$-quantile of Z:*

$$\mathrm{V@R}_\alpha(Z) := \{z \in \mathbb{R} : \mathbb{P}[Z \le z] \ge 1 - \alpha\}.$$

*This particular risk measure is widely adopted by financial institutions in order to measure the risk of portfolios. Although it satisfies the axioms (R1), (R2) and (R4), it does not satisfy, in general, the convexity axiom (R3)[35]. Thus, in general, the Value-at-Risk is not a convex risk measure. As a consequence, it turns out that risk averse stochastic programming problems with the Value-at-Risk risk measure are computationally difficult to solve (see [47]).*

[35] Note that this depends on the underlined probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For example, if we take $\Omega$ just a singleton, then $L_1(\Omega, \mathcal{F}, \mathbb{P})$ can be identified as the set of real numbers and we have that $\mathrm{V@R}_\alpha(a) = a$, for every $a \in \mathbb{R}$. Of course, this is a very artificial example.

(iii) *(Average Value-at-Risk) For every* $\alpha \in (0,1]$*, the risk measure* $\mathrm{AV@R}_\alpha : L_1(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ *is defined as:*

$$\mathrm{AV@R}_\alpha(Z) := \inf_{s\in\mathbb{R}} \left\{ s + \frac{1}{\alpha}\mathbb{E}\max\{Z-s,0\} \right\}, \forall Z \in \mathcal{Z}. \tag{2.7.3}$$

*This risk measure is closely related to the Value-at-Risk risk measure. Note that the calculation of* $\mathrm{AV@R}_\alpha(Z)$ *involves solving an optimization problem on the real line. As it is well-known (see Proposition 3.1.13), the set of* $(1-\alpha)$*-quantiles of the random variable* $Z$ *is the solution set of (2.7.3). In particular,*

$$\mathrm{V@R}_\alpha(Z) \in \operatorname*{argmin}_{s\in\mathbb{R}} \left\{ s + \frac{1}{\alpha}\mathbb{E}\max\{Z-s,0\} \right\}.$$

*It is also well-known that the* $\mathrm{AV@R}_\alpha(\cdot)$ *risk measure is a coherent risk measure (see [1]). Stochastic programming models adopting the average value-at-risk risk measure has been widely used by the stochastic programming community. One of its main features is that risk averse linear models can be solved very efficiently using this type of risk measures (see, e.g., [58]).*

(iv) *(Mean-variance) For every* $c \geq 0$*, consider the mean-variance risk measure*

$$\rho_c(Z) := \mathbb{E}Z + c\,\mathrm{Var}[Z]$$

*defined on* $L_2(\Omega, \mathcal{F}, \mathbb{P})$*. Let us consider that* $c > 0$*, otherwise we are dealing again with the expected value risk measure. In general,* $\rho_c(\cdot)$ *does not satisfy the monotonicity axiom (R1), although it satisfies axioms (R2) and (R3). It also does not satisfies axiom (R4).*

□

Note that in all examples considered above, if $Z \overset{d}{\sim} W$ with respect to $\mathbb{P}$, then $\rho(Z) = \rho(W)$. Note that $Z \overset{d}{\sim} W$ with respect to $\mathbb{P}$ if and only if $F_Z(x) = F_W(x)$, for all $x \in \mathbb{R}$, where:

$$F_Z(x) := \mathbb{P}[Z \leq x] \text{ and } F_W(x) := \mathbb{P}[W \leq x]$$

are the cumulative distribution functions of $Z$ and $W$, respectively. Note that this fact is always true, regardless of the considered probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In Section 3.1 we prove that this also true for the class of OCE risk measures. We consider below the concept of law invariant risk measures.

**Definition 2.7.5.** *(law invariant risk measures) We say that a risk measure* $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ *is law invariant, with respect to the reference probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, if* $\rho(Z) = \rho(W)$ *whenever* $Z, W \in \mathcal{Z}$ *satisfy* $Z \overset{d}{\sim} W$.

In order to consider the risk averse formulation of multistage problems, the following concept of conditional risk mappings will be useful. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$ be a sub-$\sigma$-algebra of $\mathcal{F}$, $\mathcal{Z} := L_p(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{W} := L_p(\Omega, \mathcal{G}, \mathbb{P})$, where $p \in [1, \infty)$.

**Definition 2.7.6.** *(conditional risk measures) We say that a mapping $\rho : \mathcal{Z} \to \mathcal{W}$ is a conditional risk measure if it satisfies the following property:*

*(CR1) Monotonicity: $\rho(Z) \leq \rho(W)$ for all $Z, W \in \mathcal{Z}$ such that $Z \leq W$.*

*The following axioms are also usually considered for conditional risk measures:*

*(CR2) Predictable translation equivariance: If $W \in \mathcal{W}$ and $Z \in \mathcal{Z}$, then $\rho(W + Z) = W + \rho(Z)$.*

*(CR3) Convexity: If $Z, W \in \mathcal{Z}$ and $\lambda \in [0, 1]$, then*

$$\rho(\lambda Z + (1 - \lambda)W) \leq \lambda \rho(Z) + (1 - \lambda)\rho(W).$$

*(CR4) Positive homogeneity: If $\lambda > 0$ and $Z \in \mathcal{Z}$, then $\rho(\lambda Z) = \lambda \rho(Z)$.*

*We say that a conditional risk mapping satisfying the axioms (CR1)-(CR3) is a* convex *conditional risk measure. If $\rho$ satisfies additionally the axiom (CR4) it is referred as a* coherent *conditional risk measure.*

In Chapter 4 we derive sample complexity estimates considering a risk averse multistage stochastic programming problems with conditional OCE risk measures. We show in that chapter that this class of conditional risk measures is convex. We also point that conditional OCE risk measures are not coherent, in general.

## 2.8 Miscellany

In this section we show some technical results that are used in this work. Let us begin by presenting some estimates of the covering theory in $\mathbb{R}^n$. The first result is a lemma that will be used in the derivation of the second result, which provides an upper bound estimate for the absolute constant $\rho > 0$ that appears repeatedly along the text (see, for instance, Theorem 2.1.5).

**Lemma 2.8.1.** *Consider finite constants $D > d > 0$ and $n \in \mathbb{N}$. We can cover the closed $n$-dimensional (euclidean) ball with radius $D$ with $K$ closed $n$-dimensional (euclidean) balls with radius $d$, where $K \leq (2D/d + 1)^n$.*

*Proof.* All balls considered below are closed euclidean balls of $\mathbb{R}^n$. In the following procedure, every small ball has radius $d/2$ and the big ball has radius $D$.

Place a small ball in the space with center belonging to the big ball. Suppose that we have placed $k \geq 1$ disjoints small balls in $\mathbb{R}^n$, where the center of each one belongs to the big ball[36]. If it is possible to place another small ball in the space satisfying both conditions above, proceed and make $k = k + 1$; otherwise, stop. Of course, by volume considerations (see below), the preceding algorithm stops after a finite number of iterations, say $K$. After the termination, we are in a configuration that for every point of the big ball, there exists a small ball whose distance to this point is $\leq d/2$. Indeed, if there exists a point in the big ball whose distance to each small ball is greater than $d/2$, then we can place an additional small ball with center at this point. Moreover, this small ball is disjoint to all others, which contradicts the fact that the algorithm has stopped.

Now, duplicate the radius of each one of the $K$ small balls, keeping the same center. We claim that these $K$ balls with radius $d$ cover the big one. In fact, consider an arbitrary point of the big ball. We know that there exists one small ball whose distance to this point is $\leq d/2$. By the triangular inequality, we conclude that this point belongs to the enlarged ball with radius $d$.

Finally, observe that each one of the small balls (with radius $d/2$) is contained in a ball with radius $D + d/2$. Since the small balls are disjoint, we conclude that

$$K.\operatorname{Vol}(\mathbb{B}_n)(d/2)^n \leq (D + d/2)^n \operatorname{Vol}(\mathbb{B}_n),$$

where $\operatorname{Vol}(\mathbb{B}_n)$ is the volume of the unitary euclidean ball of $\mathbb{R}^n$ and we have used that the volume of the ball with radius $r > 0$ is equal to $r^n \operatorname{Vol}(\mathbb{B}_n)$. We obtain that $K \leq (2D/d + 1)^n$, and the lemma is proved. □

**Definition 2.8.2.** *(v-net) Let $X \subseteq \mathbb{R}^n$ and $v > 0$ be given. We say that $V = \{x_1, \ldots, x_M\} \subseteq X$ is a $v$-net of $X$ if for every $x \in X$, there exists $1 \leq i \leq M$ such that: $||x_i - x|| \leq v$.*

**Proposition 2.8.3.** *Assume that $X \subseteq \mathbb{R}^n$ has finite diameter $D > 0$. For every $d \in (0, D)$, there exists a $d$-net of $X$ with $K \leq (5D/d)^n$ elements.*

*Proof.* Let $B$ be a ball with radius $D$ such that $B \supseteq X$. By Lemma 2.8.1, it is possible to cover $B$ with $\{B_i : i = 1, \ldots, K\}$, where each $B_i$ has radius $d/2$ and $K \leq (4D/d + 1)^n \leq (5D/d)^n$. Consider the following procedure. For each $i = 1, \ldots, K$, if $B_i \cap X \neq \emptyset$, select any of its elements, say $x_i$. Denote the set of selected elements by $V$. Of course, $V \subseteq X$ and $\operatorname{card} Y \leq (5D/d)^n$. We claim that $V$ is a $d$-net of $X$. In fact, take any $x \in X$. Since $X \subseteq B \subseteq \bigcup_{i=1}^{K} B_i$, there exists $i$ such that $x \in B_i$. Since $X \cap B_i \neq \emptyset$, we have selected an element $x_i \in B_i$ along the construction of

[36] We are not assuming that each small ball is contained in the big one, only its center

$V$. Since $B_i$ has radius $d/2$, the triangular inequality guarantees that $||x_i - x|| \leq d$, and the proposition is proved. □

Now we show a useful inequality involving differences of infima and the supremum of differences.

**Proposition 2.8.4.** *Let $f, g : X \to \overline{\mathbb{R}}$ be arbitrary functions, where $X$ is a nonempty set, and suppose that at least one of these functions is bounded from below. Then,*

$$\left| \inf_{x \in X} f(x) - \inf_{x \in X} g(x) \right| \leq \sup_{x \in X} |f(x) - g(x)| . \tag{2.8.1}$$

*Proof.* Suppose, without loss of generality, that $0 \leq \inf_{x \in X} f(x) - \inf_{x \in X} g(x)$. It follows that $a := \inf_{x \in X} f(x) > -\infty$, since at most one of the functions $f$ or $g$ is unbounded from below. For every $y \in X$, we have that $\inf_{x \in X} f(x) \leq f(y)$, so $\inf_{x \in X} f(x) - g(y) \leq f(y) - g(y)$, $\forall y \in X$. Taking the supremum on $y \in X$, and using that $\sup_{y \in X} -g(y) = -\inf_{y \in X} g(y)$ and $a > -\infty$, we conclude that:

$$\begin{aligned} 0 \leq \inf_{x \in X} f(x) - \inf_{y \in X} g(y) &= a - \inf_{y \in X} g(y) && (2.8.2) \\ &= \sup_{y \in X} \{a - g(y)\} && (2.8.3) \\ &\leq \sup_{y \in X} (f(y) - g(y)) && (2.8.4) \\ &= \sup_{x \in X} |f(x) - g(x)| , && (2.8.5) \end{aligned}$$

and the result is proved. □

**Remark 2.8.5.** *Observe that equation* (2.8.3) *holds, because $a > -\infty$. In fact the previous proposition is not true, in general, without supposing that at least one of the functions $f$ or $g$ is bounded from below. Consider the following counterexample: $f := g := \mathrm{Id} : \mathbb{R} \to \mathbb{R}$. Then, $\inf_{x \in \mathbb{R}} f(x) = \inf_{x \in \mathbb{R}} g(x) = \inf_{x \in \mathbb{R}} x = -\infty$, and so $\inf_{x \in \mathbb{R}} f(x) - \inf_{x \in \mathbb{R}} g(x) = (-\infty) - (-\infty) = +\infty > 0 = \sup_{x \in \mathbb{R}} |f(x) - g(x)|$.* □

**Proposition 2.8.6.** *Let $X$ be a random variable such that:*

$$M_X(s) = \mathbb{E} \exp\{sX\} < +\infty, \tag{2.8.6}$$

*for all $|s| < s_0$, where $s_0 > 0$ is a positive number. Then, $X$ has finite moments of all orders*

$$\mathbb{E}X^k < +\infty, \ \forall k \in \mathbb{N}, \tag{2.8.7}$$

*and*

$$M_X(s) = \sum_{k=0}^{\infty} \frac{s^k}{k!} \mathbb{E}X^k, \ \textit{for all } |s| < s_0. \tag{2.8.8}$$

*Proof.* See [8, Page 278]. □

Proposition 2.8.7 is a basic result about the relationship of the exponential rate of convergence of a sequence of nonnegative real numbers to 0 and the superior limit of a related sequence obtained from the convergent one. This result is often used in large deviation theory.

**Proposition 2.8.7.** *Let $\{a_n : n \in \mathbb{N}\}$ be a sequence of nonnegative real numbers. If there exists $\beta > 0$ such that*

$$\limsup_{n\to\infty} \frac{1}{n} \log a_n \leq -\beta, \tag{2.8.9}$$

*then for any $\beta' < \beta$ there exists $C = C(\beta') > 0$ finite such that*

$$a_n \leq C \exp\{-\beta' n\}, \ \forall n \in \mathbb{N}. \tag{2.8.10}$$

*Reciprocally, if $a_n \leq C \exp\{-\beta n\}$, $\forall n \in \mathbb{N}$, then*

$$\limsup_{n\to\infty} \frac{1}{n} \log a_n \leq -\beta$$

*Proof.* Take any $\beta' < \beta$. Since $-\beta < -\beta'$ and condition (2.8.9) holds, it follows that there exists $N_0 \in \mathbb{N}$ such that

$$\frac{1}{n} \log a_n \leq -\beta', \ \forall n \geq N_0.$$

So, $a_n \leq \exp\{-\beta' n\}$, for every $n \geq N_0$. Equation (2.8.10) follows by taking

$$C := \max\left\{1, a_1 \exp\{\beta'\}, \ldots, a_{N_0} \exp\{\beta' N_0\}\right\}.$$

The converse implication is elementary. □

The following example shows that, in general, it is not possible to obtain (2.8.10) with $\beta = \beta'$ from (2.8.10).

**Example 2.8.8.** *Take $a_n := n \exp\{-n\}$, for every $n \in \mathbb{N}$. Therefore, $0 \leq a_n \leq 1$, for all $n \in \mathbb{N}$ and*

$$\limsup_{n\to\infty} \frac{1}{n} \log a_n = \limsup_{n\to\infty} \left(\frac{\log n}{n} - 1\right) = -1. \tag{2.8.11}$$

*However, note that does not exist a finite $C > 0$ satisfying*

$$a_n \leq C \exp\{-n\}, \ \forall n \in \mathbb{N}.$$

□

CHAPTER 3

# Sample complexity for static problems with OCE risk measures

## 3.1 Optimized certainty equivalent risk measures

In this section we present the class of Optimized Certainty Equivalent (OCE) risk measures. This class of risk measures was extensively studied in [5].

Let us consider as given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a linear space $\mathcal{Z} := L_p(\Omega, \mathcal{F}, \mathbb{P})$, where $p \in [1, \infty)$. Here, an element $Z \in \mathcal{Z}$ represents a random loss.

**Definition 3.1.1.** *(Optimized Certainty Equivalent Risk Measures)*

*Denote by $\Phi$ the class of functions $\phi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ that are proper, l.s.c., convex and satisfy: (a) $\phi(0) = 0$, (b) $1 \in \partial\phi(0)$, and (c) $\phi$ is monotonically non-decreasing. For $\phi \in \Phi$, the OCE risk measure $\mu_\phi$ is defined as:*

$$\mu_\phi(Z) := \inf_{s\in\mathbb{R}} \{s + \mathbb{E}\phi(Z - s)\}, \text{ for } Z \in \mathcal{Z}. \tag{3.1.1}$$

□

**Remark 3.1.2.** *It may seem that we are considering a definition of OCE risk measures different from those used by other authors, like [5, 12]. But this is not the case, as we show next.*

*In [12] the author considers a class of functions $\psi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$, say $\Psi$, that are proper, l.s.c., convex and satisfy: (a) $\psi(1) = 0$, (b) $0 \in \partial\psi(1)$, (c) $\operatorname{dom}\psi \subseteq \mathbb{R}_+$. Moreover, for each $\psi \in \Psi$, the OCE risk measures $\mu_\psi$ is defined as:*

$$\mu_\psi(Z) := \inf_{s\in\mathbb{R}} \{s + \mathbb{E}\psi^*(Z - s)\}, \tag{3.1.2}$$

*where $\psi^*$ is the convex conjugate of $\psi$, see Definition 2.5.22. Defining $\Psi^* := \{\psi^* : \psi \in \Psi\}$ one can show that $\Psi^* = \Phi$; see Proposition 3.1.3. It follows that the definitions considered in this work and in [12] are equivalent.*

*Moreover, as has been pointed out in [12], the definition of OCE risk measures there is equivalent to the one given in [5].* □

The following proposition shows that the definition of Optimized Certainty Equivalent risk measures considered in this thesis is equivalent to the one considered in [12].

**Proposition 3.1.3.** *Let $\Phi$ and $\Psi$ be the sets of functions considered in Definition 3.1.1 and Remark 3.1.2, respectively. We have that:*

$$\Phi = \Psi^* := \{\psi^* : \psi \in \Psi\}. \tag{3.1.3}$$

*Proof.* $\underline{\boldsymbol{\Phi} \subseteq \boldsymbol{\Psi}^*}$ : Let $\phi \in \Phi$ be arbitrary. Since $\phi$ is proper, l.s.c. and convex, we have by Theorem 2.5.28 that $\phi = \phi^{**} = (\phi^*)^*$. Let us show that $\phi^* \in \Psi$ for concluding that $\phi \in \Psi^*$. First of all, $\phi^*$ is l.s.c. and convex by Proposition 2.5.27. Since $\phi$ is proper, we have that $\phi^*(s) > -\infty$, for all $s \in \mathbb{R}$. Moreover, since $1 \in \partial\phi(0)$, it follows from Proposition 2.5.23 that:

$$0 = 0 \cdot 1 = \phi(0) + \phi^*(1) = \phi^*(1).$$

Since we have that $\phi^{**} = \phi$, we can rewrite the previous equation as:

$$\phi^*(1) + \phi^{**}(0) = 0 = 1 \cdot 0.$$

Again by Proposition 2.5.23, we have that $0 \in \partial\phi^*(1)$. Finally, let us show that $\underline{\operatorname{dom}\phi^* \subseteq [0, +\infty)}$. Since $\operatorname{dom}\phi^*$ is convex, we have that $\operatorname{dom}\phi^* \subseteq \overline{\operatorname{dom}\phi^*} = \overline{\operatorname{ri}(\operatorname{dom}\phi^*)}$ by Proposition 2.5.13. So, let $s \in \operatorname{ri}(\operatorname{dom}\phi^*)$ be given. By Proposition 2.5.21 we have that there exists $x \in \mathbb{R}$ such that $x \in \partial\phi^*(s)$, i.e.

$$s.x = \phi^*(s) + \phi^{**}(x) = \phi^*(s) + \phi(x).$$

Then, $s \in \partial\phi(x) \subseteq [0, +\infty)$, since $\phi$ is a non-decreasing function[1]. It follows that $\operatorname{dom}\phi^* \subseteq [0, +\infty)$ and $\phi^* \in \Psi$.

$\underline{\boldsymbol{\Psi}^* \subseteq \boldsymbol{\Phi}}$ : Let $\psi \in \Psi$ be given. Let us show that $\psi^* \in \Phi$. The reasoning is similar to the previous case. By Proposition 2.5.27 we have that $\psi^*$ is l.s.c. and convex. Moreover, since $0 \in \partial\psi(1)$, it follows that:

$$0 = 1 \cdot 0 = \psi(1) + \psi^*(0) = \psi^*(0).$$

Since $\psi^{**} = \psi$, we have that:

$$\psi^{**}(1) + \psi^*(0) = 0 = 1 \cdot 0,$$

[1]Indeed, $\phi(x-1) \geq \phi(x) + s \cdot ((x-1) - x)$ implies that $s \geq \phi(x) - \phi(x-1) \geq 0$.

so $1 \in \partial\psi^*(0)$. Finally, let $x_1 < x_2$ and $t \in \operatorname{dom}\psi \subseteq [0, +\infty)$ be given. We have that:

$$t.x_1 - \psi(t) \leq t.x_2 - \psi(t),$$

for all $t \in \operatorname{dom}\psi$. Taking the supremum on $t \in \operatorname{dom}\psi$, we obtain that $\psi^*$ is a non-decreasing function. □

Given $\phi \in \Phi$, we show in Proposition 3.1.4 that $\mu_\phi(Z)$ is well-defined, whenever $Z \in L_1(\Omega, \mathcal{F}, \mathbb{P})$. Moreover if $\phi$ is Lipschitz continuous, then $\mu_\phi(Z)$ is finite. We also show that $\mu_\phi(\cdot)$ is a risk averse risk measure in $L_1(\Omega, \mathcal{F}, \mathbb{P})$ (see Definition 2.7.2). Before proceeding, let us recall that the expected value of a random variable $Z$ is not always well-defined. In fact, given a random variable $Z$, we consider its positive part $Z_+ := \max\{Z, 0\} \geq 0$ and negative part $Z_- := \max\{-Z, 0\} \geq 0$, respectively. Note that $Z = Z_+ - Z_-$ and that the expected value of each of its part is well-defined and satisfies:

$$\begin{aligned} 0 &\leq \mathbb{E}Z_+ \leq &+\infty, \text{ and} \quad (3.1.4)\\ 0 &\leq \mathbb{E}Z_- \leq &+\infty. \quad (3.1.5)\end{aligned}$$

The expected value of $Z$ is defined as:

$$\mathbb{E}Z := \mathbb{E}Z_+ - \mathbb{E}Z_- \in \overline{\mathbb{R}}, \tag{3.1.6}$$

whenever at least one of the quantities $\mathbb{E}Z_+$ and $\mathbb{E}Z_-$ is finite[2]. Note that if $\mathbb{E}\phi(Z-s)$ is well-defined, for all $s \in \mathbb{R}$, then $\mu_\phi(Z) \in \overline{\mathbb{R}}$ is also well-defined. However, in order for $\mu_\phi : \mathcal{Z} \to \overline{R}$ being a risk measure, we must also show that it is a proper function (see Definition 2.7.1).

**Proposition 3.1.4.** *Let $\phi \in \Phi$ and $Z \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ be given. Then, $\mu_\phi(Z)$ is well-defined and satisfies:*

$$-\infty < \mathbb{E}Z \leq \mu_\phi(Z) \leq \mathbb{E}\phi(Z) \leq +\infty. \tag{3.1.7}$$

*Moreover, $\mu_\phi(a) = a$, for every $a \in \mathbb{R}$, and if $\phi$ is Lipschitz continuous, then $\mathbb{E}\phi(Z)$ and $\mu_\phi(Z)$ are finite.*

*Proof.* First of all, since $\phi(0) = 0$ and $1 \in \partial\phi(0)$, we have that:

$$\phi(Z - s) \geq \phi(0) + 1 \cdot (Z - s - 0) = Z - s, \tag{3.1.8}$$

for any $s \in \mathbb{R}$. It follows that:

$$Z \leq s + \phi(Z - s), \forall s \in \mathbb{R}. \tag{3.1.9}$$

[2]Since we have already defined that $+\infty + (-\infty) = +\infty$, we could have considered that the expected value of a random variable is always a well-defined quantity. However, we prefer to disregard this convention here.

Then,

$$-\infty < \mathbb{E}Z \leq s + \mathbb{E}\phi(Z - s) \leq +\infty, \tag{3.1.10}$$

for all $s \in \mathbb{R}$. Taking the infimum on $s \in \mathbb{R}$ and considering $s = 0$, we obtain the following inequalities:

$$-\infty < \mathbb{E}Z \leq \mu_\phi(Z) = \inf_{s\in\mathbb{R}}\{s + \mathbb{E}\phi(Z - s)\} \leq \mathbb{E}\phi(Z) \leq +\infty.$$

Now, let $Z(\omega) := a$, for every $\omega \in \Omega$, where $a \in \mathbb{R}$ is given. We have that

$$\eta : s \mapsto s + \mathbb{E}\phi\left(Z - s\right) = s + \phi\left(a - s\right), \tag{3.1.11}$$

is a proper convex function that satisfies $\partial\eta(a) = \partial\eta(s)_{|s=a} = 1 - \partial\phi(a - s)_{|s=a} = 1 - \partial\phi(0)$. Since $1 \in \partial\phi(0)$, it follows that $0 \in \partial\eta(a)$, i.e., $a \in \operatorname{argmin}_{s\in\mathbb{R}} \psi(s)$. We conclude that

$$\mu_\phi(a) = \eta(a) = a + \phi(a - a) = a + \phi(0) = a. \tag{3.1.12}$$

Finally, let us suppose that $\phi$ is Lipschitz continuous, that is, there exists $L \in \mathbb{R}$ such that $|\phi(t) - \phi(s)| \leq L\,|t - s|$, for all $t, s \in \mathbb{R}$. So:

$$|\phi(Z)| = |\phi(Z) - \phi(0)| \leq L\,|Z|\,, \tag{3.1.13}$$

and we conclude that $\phi(Z) \in L_1(\Omega, \mathcal{F}, \mathbb{P})$, i.e. $\mu_\phi(Z) \leq \mathbb{E}\phi(Z) < +\infty$. $\square$

It follows from Proposition 3.1.4 that $\mu_\phi : L_1(\Omega, \mathcal{F}, \mathbb{P}) \to \overline{\mathbb{R}}$ is a proper function that satisfies:

(i) $\mu_\phi(Z) \geq \mathbb{E}Z$, for every $Z \in L_1(\Omega, \mathcal{F}, \mathbb{P})$, and

(ii) $\mu_\phi(a) = a$, for every $a \in \mathbb{R}$

for every $\phi \in \Phi$. Thus, $\mu_\phi : L_1(\Omega, \mathcal{F}, \mathbb{P}) \to \overline{\mathbb{R}}$ is a risk averse risk measure (see Definition 2.7.2). In Proposition 3.1.5 we prove that $\mu_\phi : L_1(\Omega, \mathcal{F}, \mathbb{P}) \to \overline{\mathbb{R}}$ is a convex risk measure, i.e., it satisfies axioms (R1)-(R3) below. Unless stated otherwise, we always take $\mathcal{Z} = L_1(\Omega, \mathcal{F}, \mathbb{P})$ in the remainder of this section.

**Proposition 3.1.5.** *Take any $\phi \in \Phi$. Then, $\mu_\phi : \mathcal{Z} \to \overline{\mathbb{R}}$ satisfies the following properties:*

*(R1) $\mu_\phi(Z) \leq \mu_\phi(W)$, for every $Z, W \in \mathcal{Z}$, whenever $Z \leq W$.*

*(R2) $\mu_\phi(Z + a) = \mu_\phi(Z) + a$, for every $Z \in \mathcal{Z}$ and $a \in \mathbb{R}$.*

*(R3) $\mu_\phi(\lambda Z + (1 - \lambda)W) \leq \lambda\mu_\phi(Z) + (1 - \lambda)\mu_\phi(W)$, for every $Z, W \in \mathcal{Z}$ and $\lambda \in [0, 1]$.*

*Proof.* (R1): Take any $Z, W \in \mathcal{Z}$ satisfying $Z \leq W$. For any $s \in \mathbb{R}$, we have that $\mathbb{E}\phi(Z-s) \leq \mathbb{E}\phi(W-s)$, since $\phi(\cdot)$ is a non-decreasing function. It follows that

$$\mu_\phi(Z) \leq s + \mathbb{E}\phi(Z-s) \leq s + \mathbb{E}\phi(W-s), \tag{3.1.14}$$

for all $s \in \mathbb{R}$. Taking the infimum in $s \in \mathbb{R}$ above, we conclude that $\mu_\phi(Z) \leq \mu_\phi(W)$, which proves (R1).

(R2): Take any $Z \in \mathcal{Z}$ and $a \in \mathbb{R}$. Note that

$$\begin{aligned} \mu_\phi(Z+a) &= \inf_{s\in\mathbb{R}} \{s + \mathbb{E}\phi(Z+a-s)\} & (3.1.15)\\ &= \inf_{s\in\mathbb{R}} \{a + (s-a) + \mathbb{E}\phi(Z-(s-a))\} & (3.1.16)\\ &= a + \inf_{s\in\mathbb{R}} \{s - a + \mathbb{E}\phi(Z-(s-a))\} & (3.1.17)\\ &= a + \mu_\phi(Z), & (3.1.18) \end{aligned}$$

which proves (R2).

(R3): Take any $Z, W \in \mathcal{Z}$ and $\lambda \in [0,1]$. Suppose, without loss of generality, that $\mu_\phi(Z)$ and $\mu_\phi(W)$ are finite. So, let $s_1, s_2 \in \mathbb{R}$ be such that $\mathbb{E}\phi(Z-s_1)$ and $\mathbb{E}\phi(W-s_2)$ are finite. It follows that

$$\begin{aligned} &\lambda\left[s_1 + \mathbb{E}\phi(Z-s_1)\right] + (1-\lambda)\left[s_2 + \mathbb{E}\phi(W-s_2)\right] = \\ &\quad = \lambda s_1 + (1-\lambda)s_2 + \mathbb{E}\left[\lambda\phi(Z-s_1) + (1-\lambda)\phi(W-s_2)\right] \\ &\qquad \geq \lambda s_1 + (1-\lambda)s_2 + \mathbb{E}\phi\left(\lambda Z + (1-\lambda)W - [\lambda s_1 + (1-\lambda)s_2]\right) \\ &\qquad\quad \geq \mu_\phi\left(\lambda Z + (1-\lambda)W\right). \end{aligned}$$

Taking the infimum in $s_1, s_2 \in \mathbb{R}$ above, it follows that $\lambda\mu_\phi(Z) + (1-\lambda)\mu_\phi(W) \geq \mu_\phi(\lambda Z + (1-\lambda)W)$. This completes the proof of the proposition. □

It is not true, in general, that an OCE risk measure is *coherent*. In fact, taking $\phi(s) = \exp(s) - 1$, for all $s \in \mathbb{R}$, we show in Example 3.1.18 that

$$\mu_\phi(Z) = \log\left(\mathbb{E}\exp\{Z\}\right) = \log\left(M_Z(1)\right), \tag{3.1.19}$$

for any $Z \in \mathcal{Z}$. As usual, we denote by $M_Z(\cdot)$ the moment generating function of $Z$. So, let $Z \overset{d}{\sim} \text{Gaussian}(0,1)$[3] and $\lambda > 0$ be given. Noting that $\lambda Z \overset{d}{\sim} \text{Gaussian}(0,\lambda^2)$ and $M_{\text{Gaussian}(0,\sigma^2)}(s) = \exp\{\sigma^2 s^2/2\}$, for all $s \in \mathbb{R}$, we obtain that

$$\mu_\phi(\lambda Z) = \log\left(\exp\{\lambda^2/2\}\right) = \lambda^2/2 \neq \lambda/2 = \lambda\mu_\phi(Z), \tag{3.1.20}$$

[3] Of course, we can consider a sufficiently rich probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that there exists a random variable $Z \in \mathcal{Z} = L_1(\Omega, \mathcal{F}, \mathbb{P})$ that has standard Gaussian distribution. Indeed, take $\Omega := [0,1]$, $\mathcal{F} := \mathcal{B}([0,1])$ and $\mathbb{P} =$ "the Lebesgue measure on $[0,1]$". For every $F \in \mathcal{D}_1$, consider $F^{-1}(\omega) := \inf\{z \in \mathbb{R} : F(z) \leq \omega\}$, for every $\omega \in [0,1]$. It is elementary to show that the cumulative distribution function of $Z := F^{-1}$ is $F_Z = F$. Since $F \in \mathcal{D}_1$, it follows that $Z \in \mathcal{Z}$. Now, taking $F(z) := \frac{1}{\sqrt{2\pi}} \int_{(-\infty,z]} \exp\left\{-\frac{x^2}{2}\right\} dx$, we obtain that $Z \overset{d}{\sim} \text{Gaussian}(0,1)$.

for any positive $\lambda \neq 1$. Thus, we conclude that $\mu_\phi(\cdot)$ is not positively homogeneous, in particular, it is not a coherent risk measure.

Take any $\phi \in \Phi$ and consider $\mathcal{Z} = L_1(\Omega, \mathcal{F}, \mathbb{P})$, where $(\Omega, \mathcal{F}, \mathbb{P})$ is an arbitrary probability space. We have already shown that $\mu_\phi : \mathcal{Z} \to$ is a well-defined risk measure. Changing the probability space $\mathcal{Z}$ to another one, we obtain a different risk measure. We have already shown in Proposition 3.1.5 that regardless the reference probability space that we consider, $\mu_\phi : \mathcal{Z} \to \overline{\mathbb{R}}$ is always a convex risk measure. In Proposition 3.1.6 we show that this is also true regarding the law invariance property of $\mu_\phi : \mathcal{Z} := L_1(\Omega, \mathcal{F}, \mathbb{P}) \to \overline{\mathbb{R}}$ in $\mathcal{Z}$. It is worth mentioning that a risk measure can be law invariant with respect to some reference probability, but could fail to satisfy this property when one considers another reference probability space (see [73, Example 6.48]).

**Proposition 3.1.6.** *Take any probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{Z} := L_1(\Omega, \mathcal{F}, \mathbb{P})$. Let $\phi \in \Phi$ be arbitrary. We have that*

$$\mu_\phi(Z) = \mu_\phi(W), \tag{3.1.21}$$

*for all $Z, W \in \mathcal{Z}$ such that $Z \overset{d}{\sim} W$.*

*Proof.* Let $Z, W \in \mathcal{Z}$ be such that $Z \overset{d}{\sim} W$. Take any $s \in \mathbb{R}$. We already know by Proposition 3.1.4 that $\mathbb{E}\phi(Z - s)$ is well-defined, for any $Z \in \mathcal{Z}$. Since $Z \overset{d}{\sim} W$, it follows that $\phi(Z - s) \overset{d}{\sim} \phi(W - s)$. Thus,

$$s + \mathbb{E}\phi(Z - s) = s + \mathbb{E}\phi(W - s), \tag{3.1.22}$$

for every $s \in \mathbb{R}$. We conclude that $\mu_\phi(Z) = \mu_\phi(W)$, which proves the proposition. □

Before presenting some examples of OCE risk measures, we show the following result, to be used in item (c) of Example 3.1.8.

**Proposition 3.1.7.** *The class of functions $\Phi$ is a convex set, i.e. $\forall \phi_1, \phi_2 \in \Phi$ and $0 \leq \lambda \leq 1$, we have that $(1 - \lambda)\phi_1 + \lambda\phi_2 \in \Phi$.*

*Proof.* Let us verify that $\psi := (1 - \lambda)\phi_1 + \lambda\phi_2$ is l.s.c., convex and satisfies items (a)-(c) of Definition 3.1.1. First of all, note that:

$$\psi(0) = (1 - \lambda)\phi_1(0) + \lambda\phi_2(0) = 0,$$

which shows item (a). Moreover, since $\lambda$ and $1 - \lambda$ are non-negative numbers and $\phi_1$ and $\phi_2$ are monotonically non-decreasing, it follows that $\psi$ is monotonically non-decreasing, which shows item (c). We also have that:

$$(1 - \lambda)\partial\phi_1(x) + \lambda\partial\phi_2(x) \subseteq \partial\psi(x),$$

for any $x \in \mathbb{R}$. Indeed, if $\partial\phi_1(x) = \emptyset$ or $\partial\phi_2(x) = \emptyset$, the claim is trivially true. So, let $s_i \in \partial\phi_i(x)$ be given ($i = 1, 2$). We have that:

$$\begin{aligned}
\psi(y) &= (1-\lambda)\phi_1(y) + \lambda\phi_2(y) && (3.1.23)\\
&\geq (1-\lambda)[\phi_1(x) + s_1(y-x)] + \lambda[\phi_2(x) + s_2(y-x)] && (3.1.24)\\
&= (1-\lambda)\phi_1(x) + \lambda\phi_2(x) + [(1-\lambda)s_1 + \lambda s_2](y-x) && (3.1.25)\\
&= \psi(x) + [(1-\lambda)s_1 + \lambda s_2](y-x), && (3.1.26)
\end{aligned}$$

for all $y \in \mathbb{R}$. We conclude that $(1-\lambda)s_1 + \lambda s_2 \in \partial\psi(x)$. Applying the result for $s_1 = 1 = s_2$ and $x = 0$, we obtain that $1 \in \partial\psi(0)$, which shows item (b). The proof that $\psi$ is l.s.c. and convex is elementary. □

Now, let us consider some examples of OCE risk measures.

**Example 3.1.8.** *In all the examples below, $\mu_\phi$ ($\phi \in \Phi$) is defined on $\mathcal{Z} := L_1(\Omega, \mathcal{F}, \mathbb{P})$, where $(\Omega, \mathcal{F}, \mathbb{P})$ is a given probability space. That is, all random losses $Z$ are integrable random variables.*

*(a) (Expected value operator) Consider $\phi := \mathrm{Id} : \mathbb{R} \to \mathbb{R}$, i.e. $\phi(z) = z$, for all $z \in \mathbb{R}$. It is straightforward to verify that $\phi \in \Phi$. Let $Z \in \mathcal{Z}$ and $s \in \mathbb{R}$ be given, we have that:*

$$s + \mathbb{E}\phi(Z - s) = \mathbb{E}Z, \tag{3.1.27}$$

*i.e. $\mu_\phi(Z) = \mathbb{E}Z$. Therefore the expected value operator is an example of OCE risk measure.*

*(b) (Average Value-at-Risk) For $\alpha \in [0, 1)$ given, consider $\phi(z) := \frac{1}{1-\alpha}\max\{z, 0\}$, for all $z \in \mathbb{R}$. It is straightforward to verify that $\phi \in \Phi$. Observe that:*

$$1 \in \partial\phi(0) = [0, 1/(1-\alpha)]\,. \tag{3.1.28}$$

*The associated risk measure:*

$$\mathrm{AV@R}_{1-\alpha}(Z) := \inf_{s \in \mathbb{R}} \left\{ s + \frac{1}{1-\alpha}\mathbb{E}\max\{Z - s, 0\} \right\}, \tag{3.1.29}$$

*is known as the Average Value-at-Risk risk measure.*

*(c) (Convex combination of OCE risk measure with the expected value operator) Let $\phi \in \Phi$ be given. We have that:*

$$\mu(Z) := (1-\lambda)\mathbb{E}Z + \lambda\mu_\phi(Z)$$

*is also an OCE risk measure. We claim that $\mu = \mu_\psi$, where $\psi := (1-\lambda)\,\mathrm{Id} + \lambda\phi$. Before showing our claim, observe that $\psi \in \Phi$ by Proposition 3.1.7 which shows that $\mu_\psi$ is an OCE risk measure. For every $s \in \mathbb{R}$, we have that:*

$$\mathbb{E}Z = s + \mathbb{E}(Z - s). \tag{3.1.30}$$

*It follows that:*

$$\begin{aligned}
\mu(Z) &= (1-\lambda)\mathbb{E}Z + \lambda \inf_{s\in\mathbb{R}} \{s + \mathbb{E}\phi(Z-s)\} && (3.1.31)\\
&= \inf_{s\in\mathbb{R}} \{(1-\lambda)[s+\mathbb{E}(Z-s)] + \lambda[s+\mathbb{E}\phi(Z-s)]\} && (3.1.32)\\
&= \inf_{s\in\mathbb{R}} \{s + \mathbb{E}((1-\lambda)\operatorname{Id} + \lambda\phi)(Z-s)\} && (3.1.33)\\
&= \mu_\psi(Z), && (3.1.34)
\end{aligned}$$

*and the claim is proved. One particular example of such risk measure is the convex combination of the expected value operator and the Average Value-at-Risk.* □

Given $\phi \in \Phi$, consider the set of all subgradients of $\phi$ (see Definition 2.5.37):

$$\mathcal{I}_\phi = \bigcup_{z\in\operatorname{dom}\phi} \partial\phi(z). \quad (3.1.35)$$

By Corollary 2.5.36, we have that $\mathcal{I}_\phi \subseteq \mathbb{R}$ is a nonempty interval. By Proposition 2.5.38, we have that its extreme points $l(\phi) := \inf \mathcal{I}_\phi$ and $L(\phi) := \sup \mathcal{I}_\phi$ satisfy:

$$0 \leq l(\phi) \leq L(\phi) \leq +\infty.$$

Moreover, $\phi$ is a Lipschitz continuous function if and only if $L(\phi) < +\infty$, in which case, $L(\phi)$ is a Lipschitz constant of $\phi$. We show in Proposition 3.1.9 that if $L(\phi) < \infty$, then $\mu_\phi : \mathcal{Z} \to \mathbb{R}$ is $L(\phi)$-Lipschitz continuous. This result will be used often in the sequel.

**Proposition 3.1.9.** *Take any $\phi \in \Phi$ such that $L(\phi) < \infty$. We have that $\mu_\phi : \mathcal{Z} \to \mathbb{R}$ is $L(\phi)$-Lipschitz continuous in $\mathcal{Z}$ considering the $L_1$-norm:*

$$\|Z\|_1 := \mathbb{E}|Z|, \ \forall Z \in \mathcal{Z}. \quad (3.1.36)$$

*Proof.* Let $\phi \in \Phi$ be such that $L(\phi) < \infty$ and take any $Z, W \in \mathcal{Z}$. We have that

$$\begin{aligned}
|\mu_\phi(Z) - \mu_\phi(W)| &= \left|\inf_{s\in\mathbb{R}} \{s + \mathbb{E}\phi(Z-s)\} - \inf_{s\in\mathbb{R}} \{s + \mathbb{E}\phi(W-s)\}\right| && (3.1.37)\\
&\leq \sup_{s\in\mathbb{R}} |(s + \mathbb{E}\phi(Z-s)) - (s + \mathbb{E}\phi(W-s))| && (3.1.38)\\
&\leq \sup_{s\in\mathbb{R}} \mathbb{E}|\phi(Z-s) - \phi(W-s)| && (3.1.39)\\
&\leq \sup_{s\in\mathbb{R}} L(\phi)\mathbb{E}|(Z-s) - (W-s)| && (3.1.40)\\
&= L(\phi)\mathbb{E}|Z-W| && (3.1.41)\\
&= L(\phi)\|Z-W\|_1. && (3.1.42)
\end{aligned}$$

It is elementary to verify the validity of each equation above, let us just mention that we used Proposition 2.8.4 in (3.1.38). This completes the proof of the result. □

Let $Z$ be an integrable random variable. We will establish some properties regarding the optimal set and the optimal value of the optimization problem on $\mathbb{R}$:

$$\mu_\phi(Z) := \inf_{s\in\mathbb{R}} \left\{\eta(s) := s + \mathbb{E}\phi(Z-s)\right\}. \tag{3.1.43}$$

In the next lemma, we obtain bounds for the right and left derivatives of $\phi$ on $\operatorname{dom}\phi$.

**Lemma 3.1.10.** *Let $\phi \in \Phi$ be given. For all $z \in \operatorname{dom}\phi$ such that $\partial\phi(z) \neq \emptyset$, we have that:*

$$l(\phi) \le -\phi'(z;-1) \le \phi'(z;1) \le L(\phi). \tag{3.1.44}$$

*Proof.* Let $z \in \operatorname{dom}\phi$ be such that $\partial\phi(z) \neq \emptyset$. By Corollary 2.5.33, we have that:

$$\partial\phi(z) = [-\phi'(z;-1), \phi(z;1)] \cap \mathbb{R}. \tag{3.1.45}$$

Moreover, since $\partial\phi(z) \neq \emptyset$, we have that $\inf\partial\phi(z) = -\phi'(z;-1)$ and $\sup\partial\phi(z) = \phi'(z;1)$. We also have that $\partial\phi(z) \subseteq \mathcal{I}_\phi$, so:

$$l(\phi) = \inf\mathcal{I}_\phi \le \inf\partial\phi(z) = -\phi'(z;-1) \le \phi'(z;1) = \sup\partial\phi(z) \le \sup\mathcal{I}_\phi = L(\phi), \tag{3.1.46}$$

and the lemma is proved. ☐

In the next proposition we show that when any of these conditions hold:

(i) $L(\phi) = 1$,

(ii) $l(\phi) = 1$ and $\phi$ is Lipschitz continuous[4],

the OCE risk measure $\mu_\phi$ is just the expected value operator on $\mathcal{Z}$. By Proposition 3.1.11, we also have that $l(\phi) \le 1 \le L(\phi)$, for $\phi \in \Phi$. So, the interesting situation to analyze for OCE risk measures with $\phi$ Lipschitz continuous occurs when $l(\phi) < 1 < L(\phi)$.

**Proposition 3.1.11.** *Take $Z \in \mathcal{Z}$ and $\phi \in \Phi$ such that $L(\phi) < +\infty$. Let us consider the objective function $\eta(s) := s + \mathbb{E}\phi(Z-s)$, for all $s \in \mathbb{R}$. The following assertions hold:*

*(a) If $l(\phi) = 1$, then $\mu_\phi(Z) = \lim_{s\to+\infty} \eta(s) = \mathbb{E}Z$.*

*(b) If $L(\phi) = 1$, then $\mu_\phi(Z) = \lim_{s\to-\infty} \eta(s) = \mathbb{E}Z$.*

[4] As pointed out previously, $L(\phi) < +\infty$ if and only if $\phi$ is Lipschitz continuous. So, item (i) automatically implies that $\phi$ is Lipschitz continuous.

*Proof.* Since $Z \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ and $\phi$ is Lipschitz continuous, it follows from Proposition 2.5.43 that:

$$\mathbb{E}\phi(Z - s) \in \mathbb{R}, \tag{3.1.47}$$

for all $t \in \mathbb{R}$. By Proposition 2.5.42, we conclude that $\eta$ is a finite-valued convex function that satisfies:

$$\partial\eta(s) = [1 - \mathbb{E}\phi'(Z - s; 1), 1 + \mathbb{E}\phi'(Z - s; -1)], \tag{3.1.48}$$

for all $s \in \mathbb{R}$. Since $\phi$ is Lipschitz continuous, we have that $\operatorname{dom}\phi = \mathbb{R}$. Given $s \in \mathbb{R}$, we have by Lemma 3.1.10 that:

$$l(\phi) \le -\phi'(z - s; -1) \le \phi'(z - s; 1) \le L(\phi), \tag{3.1.49}$$

for all $z \in \mathbb{R}$. Therefore we obtain the following bounds:

$$l(\phi) \le -\phi'(Z - s; -1) \le \phi'(Z - s; 1) \le L(\phi) \tag{3.1.50}$$

for all $s \in \mathbb{R}$. Taking the expected value, we obtain that:

$$l(\phi) \le -\mathbb{E}\phi'(z - s; -1) \le \mathbb{E}\phi'(z - s; 1) \le L(\phi), \tag{3.1.51}$$

for all $s \in \mathbb{R}$. It follows that:

$$\begin{array}{rcccl} -\eta'(s; -1) & = & 1 - \mathbb{E}\phi'(Z - s; 1) & \ge & 1 - L(\phi), \\ \eta'(s; 1) & = & 1 + \mathbb{E}\phi'(Z - s; -1) & \le & 1 - l(\phi), \end{array} \tag{3.1.52}$$

for all $s \in \mathbb{R}$.

Since $\eta$ is finite-valued, we have that $\operatorname{int}(\operatorname{dom}\eta) = \mathbb{R}$. Applying Theorem 2.5.40 to the convex function $\eta$, we obtain that:

$$\eta(s) - \eta(t) = \int_t^s \eta'(u; 1)du = \int_t^s -\eta'(u; -1)du, \tag{3.1.53}$$

for all $t < s$.

Now, suppose that $l(\phi) = 1$. We obtain that $\eta'(u; 1) \le 1 - l(\phi) = 0$, for all $u \in \mathbb{R}$. By equation (3.1.53) we conclude that $\eta(s) \le \eta(t)$, for $s > t$. It follows that:

$$\mu_\phi(Z) := \inf_{s \in \mathbb{R}} \eta(s) = \lim_{s \to \infty} \eta(s). \tag{3.1.54}$$

Moreover, observe that for every $s \in \partial\phi(z)$ with $z < 0$, we have that $1 = l(\phi) \le s \le 1 \in \partial\phi(0)$, i.e. $s = 1$. We conclude that $\phi'(z) = 1$ for all $z < 0$. Since $\phi(0) = 0$, it follows from equation (3.1.53) that $\phi(z) = z$, for all $z < 0$. Let $s \ge 0$ be arbitrary. We have that:

$$\begin{aligned} \eta(s) = s + \mathbb{E}\phi(Z - s) &= \mathbb{E}\left[(\phi(Z - s) + s)\mathbb{1}_{\{Z<s\}}\right] + \mathbb{E}\left[(\phi(Z - s) + s)\mathbb{1}_{\{Z\ge s\}}\right] \\ &= \mathbb{E}\left[Z\mathbb{1}_{\{Z<s\}}\right] + r(s), \end{aligned} \tag{3.1.55}$$

where:

$$\begin{aligned} 0 \leq r(s) &:= \mathbb{E}\left[(\phi(Z-s)+s)\mathbb{1}_{\{Z\geq s\}}\right] \\ &\leq \mathbb{E}\left[(L(\phi)(Z-s)+s)\mathbb{1}_{\{Z\geq s\}}\right] \\ &\leq L(\phi)\mathbb{E}\left[Z\mathbb{1}_{\{Z\geq s\}}\right], \end{aligned}$$

since $s \geq 0$, $L(\phi) \geq 1$, $Z \geq s$ and $\phi(z) \leq L(\phi)z$, for $z \geq 0$. Since $Z$ is integrable, we have that:

$$\lim_{s\to+\infty} \mathbb{E}Z\mathbb{1}_{\{Z\geq s\}} = 0.$$

We conclude that $r(s)$ converges to zero, as $s \to +\infty$, hence $\lim_{s\to+\infty} \eta(s) = \mathbb{E}Z$, which proves item (a).

Now, suppose that $L(\phi) = 1$. We obtain that $-\eta'(u; -1) \geq 1 - L(\phi) \geq 0$, for all $u \in \mathbb{R}$. It follows that $\eta(s) \geq \eta(t)$, for all $s > t$. So,

$$\mu_\phi(Z) := \inf_{s\in\mathbb{R}} \eta(s) = \lim_{s\to-\infty} \eta(s). \tag{3.1.56}$$

Moreover, observe that for all $s \in \partial\phi(z)$ with $z > 0$, we have that $\partial\phi(0) \ni 1 \leq s \leq L(\phi) = 1$, i.e. $s = 1$. This shows that $\phi'(z) = 1$, for $z > 0$. Since $\phi(0) = 0$, it follows that $\phi(z) = z$, for all $z > 0$. Let $s \leq 0$ be arbitrary. We have that:

$$\begin{aligned} \eta(s) = s + \mathbb{E}\phi(Z-s) &= \mathbb{E}\left[(\phi(Z-s)+s)\mathbb{1}_{\{Z>s\}}\right] + \mathbb{E}\left[(\phi(Z-s)+s)\mathbb{1}_{\{Z\leq s\}}\right] \\ &= \mathbb{E}\left[Z\mathbb{1}_{\{Z>s\}}\right] + R(s) \\ &\leq \mathbb{E}\left[Z\mathbb{1}_{\{Z>s\}}\right], \end{aligned} \tag{3.1.57}$$

since $R(s) := \mathbb{E}\left[(\phi(Z-s)+s)\mathbb{1}_{\{Z\leq s\}}\right] \leq 0$. In fact, for $Z \leq s \leq 0$, we have that $\phi(Z-s)+s \leq \phi(0)+s \leq 0$. Letting $s \to -\infty$, we obtain that

$$(\mathbb{E}Z \leq) \lim_{s\to-\infty} \eta(s) \leq \lim_{s\to-\infty} \mathbb{E}\left[Z\mathbb{1}_{\{Z>s\}}\right] = \mathbb{E}Z,$$

where the first inequality follows from Proposition 3.1.4. So, the result is proved. □

**Remark 3.1.12.** *It follows from the previous proposition that* $\mathrm{AV@R}_1(Z) = \mathbb{E}Z$*, for* $Z \in L_1(\Omega, \mathcal{F}, \mathbb{P})$*. In fact, in this case we have that* $\phi(z) = \max\{z, 0\}$*, so* $L(\phi) = 1$*. This particular result is well-known.* □

Now, we begin to analyze the situation $l(\phi) < 1 < L(\phi) < +\infty$. Before proceeding, we show a well-known result on the AV@R risk measure. In Section 2.3 we recall the definition of an $\alpha$-quantile of a random variable $Z$. When $\alpha \in (0, 1)$ the set of $\alpha$-quantiles of $Z$ is a nonempty compact interval whose extreme points are denoted as $q^-_\alpha(Z) \leq q^+_\alpha(Z)$ (see Proposition 2.3.2).

**Proposition 3.1.13.** *Let be given $Z \in \mathcal{Z}$ and $0 < \alpha < 1$. For the* $\mathrm{AV@R}_{1-\alpha}(\cdot)$ *risk measure, the set of optimal solutions of the optimization problem:*

$$\min_{s\in\mathbb{R}} \left\{ s + \frac{1}{1-\alpha}\mathbb{E}\max\{Z - s, 0\} \right\} \tag{3.1.58}$$

*is the set of $\alpha$-quantiles of $Z$.*

*Proof.* Let us denote the objective function of the optimization problem by $\eta(s) := s + \mathbb{E}\phi(Z - s)$, where:

$$\phi(z) = \frac{1}{1-\alpha}\max\{z, 0\}, \forall z \in \mathbb{R}. \tag{3.1.59}$$

The right and left derivatives of $\phi$ are:

$$\phi'(z; 1) = \begin{cases} 0, & \text{for } z < 0, \\ 1/(1-\alpha), & \text{for } z \geq 0, \end{cases} \tag{3.1.60}$$

and:

$$-\phi'(z; -1) = \begin{cases} 0, & \text{for } z \leq 0, \\ 1/(1-\alpha), & \text{for } z > 0, \end{cases}, \tag{3.1.61}$$

respectively. It follows from Proposition 2.5.42 that:

$$-\eta'(s; -1) = 1 - \mathbb{E}\phi'(Z - s; 1) = 1 - \frac{1}{1-\alpha}\mathbb{E}\mathbb{1}_{\{Z \geq s\}} = 1 - \frac{1}{1-\alpha}\mathbb{P}[Z \geq s], \tag{3.1.62}$$

$$\eta'(s; 1) = 1 + \mathbb{E}\phi'(Z - s; -1) = 1 - \frac{1}{1-\alpha}\mathbb{E}\mathbb{1}_{\{Z > s\}} = 1 - \frac{1}{1-\alpha}\mathbb{P}[Z > s]. \tag{3.1.63}$$

We also have that $\bar{s} \in \operatorname{argmin}_{s\in\mathbb{R}} \eta(s)$ if and only if $0 \in \partial\eta(\bar{s}) = [-\eta'(\bar{s}; -1), \eta'(\bar{s}; 1)]$, i.e. $-\eta'(s; -1) \leq 0 \leq \eta'(s; 1)$. Hence $\bar{s}$ must be such that:

$$\mathbb{P}[Z \geq \bar{s}] \geq 1 - \alpha, \text{ and} \tag{3.1.64}$$

$$\mathbb{P}[Z \leq \bar{s}] \geq \alpha. \tag{3.1.65}$$

We conclude that $\bar{s} \in \operatorname{argmin}_{s\in\mathbb{R}} \eta(s)$ if and only if $\bar{s}$ is an $\alpha$-quantile of $Z$. □

**Remark 3.1.14.** *Let us point out that for the* $\mathrm{AV@R}_{1-\alpha}(\cdot)$ *risk measure* ($\alpha \in (0, 1)$) *Proposition 2.5.42 can be applied supposing just that $\mathbb{E}Z_+ < +\infty$. This condition is weaker than assuming that $Z \in L_1(\Omega, \mathcal{F}, \mathbb{P})$. In fact, given $s \in \mathbb{R}$, we have that:*

$$0 \leq \max\{Z - s, 0\} \leq Z_+ + |s|. \tag{3.1.66}$$

*Therefore $\eta(\cdot)$ is finite-valued when $\mathbb{E}Z_+ < +\infty$.* □

We will extend Proposition 3.1.13 to general OCE risk measures $\mu_\phi$ satisfying $l(\phi) < 1 < L(\phi) < +\infty$. In fact, in Proposition 3.1.17 we show that the solution set of problem:

$$\min_{s\in\mathbb{R}} \{s + \mathbb{E}\phi(Z - s)\} \tag{3.1.67}$$

is a nonempty interval that is contained in a bounded interval whose extreme points depend on particular quantiles of the random variable $Z$, and on particular finite numbers $\underline{z} \le 0 \le \overline{z}$ that depend only on $\phi$. The following lemma will be useful.

**Lemma 3.1.15.** *Take $Z \in \mathcal{Z}$ and $\phi \in \Phi$ satisfying $l(\phi) < 1 < L(\phi) < +\infty$. The following assertions hold:*

*(i) There exist real constants $\underline{z} \le 0 \le \overline{z}$ depending only on $\phi$ such that:*

$$\phi'(z;1) \;\ge\; \frac{1+L(\phi)}{2}, \text{ for all } z \ge \overline{z}, \tag{3.1.68}$$

$$-\phi'(z;-1) \;\le\; \frac{1+l(\phi)}{2}, \text{ for all } z \le \underline{z}. \tag{3.1.69}$$

*(ii) If $s \le q^+_\alpha(Z) - \overline{z}$, then*

$$\mathbb{E}\left[\phi'(Z-s;1)\right] \ge l(\phi)\alpha + \frac{1+L(\phi)}{2}(1-\alpha). \tag{3.1.70}$$

*(iii) If $s \ge q^-_\alpha(Z) - \underline{z}$, then*

$$\mathbb{E}\left[-\phi'(Z-s;-1)\right] \le L(\phi) + \frac{1+l(\phi)-2L(\phi)}{2}\alpha. \tag{3.1.71}$$

*Proof.* Firstly, we show item (i). Since $L(\phi) < +\infty$, we have that $\phi$ is finite-valued and $\phi'(\cdot;1)$ is a monotonically non-decreasing finite-valued function. Moreover, we have that:

$$L(\phi) = \sup \bigcup_{z\in\mathbb{R}} \partial\phi(z) = \sup \bigcup_{z\in\mathbb{R}} [-\phi'(z;-1), \phi'(z;1)] = \sup_{z\in\mathbb{R}} \phi'(z;1).$$

Since $L(\phi) > 1$, it follows that $(1 + L(\phi))/2 < L(\phi)$. Thus there exists $\overline{z} \in \mathbb{R}$ such that $\phi'(\overline{z};1) \ge (1 + L(\phi))/2$. Therefore, if $z \ge \overline{z}$, we have that $\phi'(z;1) \ge \phi'(\overline{z};1) \ge (1 + L(\phi))/2$. Observe also that if $z < 0$, then:

$$\phi'(z;1) \le -\phi'(0;-1) \le 1 < (1 + L(\phi))/2 \le \phi'(\overline{z};1),$$

which implies that $\bar{z} \ge 0$. The proof that there exists $\underline{z} \le 0$ such that $-\phi'(z;-1) \le (1 + l(\phi))/2$ for all $z \le \underline{z}$ is similar. Just note that $-\phi'(\cdot;-1)$ is monotonically non-decreasing; $-\phi'(z;-1) \ge 1$, for all $z > 0$; $l(\phi) < (1 + l(\phi))/2 < 1$ and $l(\phi) = \inf_{z\in\mathbb{R}} -\phi'(z;-1)$. We have shown item (i).

Now, let us show item (ii). Denote by $\mathbb{P}_Z$ the probability measure induced by $\mathbb{P}$ and $Z$ on $\mathbb{R}$, that is:

$$\mathbb{P}_Z(B) := \mathbb{P}\left[Z \in B\right], \ \forall B \in \mathbb{B}(\mathbb{R}). \tag{3.1.72}$$

Let $s \leq q_\alpha^+(Z) - \overline{z}$ be given. We have that:

$$\begin{aligned}
\mathbb{E}\left[\phi'(Z-s;1)\right] &= \int_{\mathbb{R}} \phi'(z-s;1)\mathbb{P}_Z(dz) \\
&= \int_{(-\infty,\overline{z}+s)} \phi'(z-s;1)\mathbb{P}_Z(dz) + \int_{[\overline{z}+s,\infty)} \phi'(z-s;1)\mathbb{P}_Z(dz) \\
&\geq \int_{(-\infty,\overline{z}+s)} l(\phi)\mathbb{P}_Z(dz) + \int_{[\overline{z}+s,\infty)} \frac{1+L(\phi)}{2}\mathbb{P}_Z(dz) \\
&= l(\phi)\left(1-\mathbb{P}[Z \geq \overline{z}+s]\right) + \frac{1+L(\phi)}{2}\mathbb{P}[Z \geq \overline{z}+s],
\end{aligned} \tag{3.1.73}$$

where the inequality follows from $\phi'(\cdot;1) \geq l(\phi)$ and $\phi'(z-s;1) \geq (1+L(\phi)/2$, for $z \geq \overline{z}+s$. Since $\overline{z}+s \leq q_\alpha^+(Z)$, we have that:

$$\mathbb{P}[Z \geq \overline{z}+s] \geq \mathbb{P}[Z \geq q_\alpha^+(Z)] \geq 1-\alpha. \tag{3.1.74}$$

The affine function $\theta \in [1-\alpha,+\infty) \mapsto l(\phi)(1-\theta) + (1+L(\phi))\theta/2$ attains its minimum value at $\theta = 1-\alpha$, since $l(\phi) < (1+L(\phi))/2$. Therefore, we obtain that:

$$\mathbb{E}\left[\phi'(Z-s;1)\right] \geq l(\phi)\alpha + \frac{1+L(\phi)}{2}(1-\alpha), \tag{3.1.75}$$

which shows item (ii).

Now, let us show item (iii). Given $s \geq q_\alpha^-(Z) - \underline{z}$, we have that:

$$\begin{aligned}
\mathbb{E}\left[-\phi'(Z-s;-1)\right] &= \int_{\mathbb{R}} -\phi'(z-s;-1)\mathbb{P}_Z(dz) \\
&= \int_{(-\infty,\underline{z}+s)} -\phi'(z-s;1)\mathbb{P}_Z(dz) + \int_{[\underline{z}+s,\infty)} -\phi'(z-s;1)\mathbb{P}_Z(dz) \\
&\leq \int_{(-\infty,\underline{z}+s]} \frac{1+l(\phi)}{2}\mathbb{P}_Z(dz) + \int_{(\underline{z}+s,\infty)} L(\phi)\mathbb{P}_Z(dz) \\
&= \frac{1+l(\phi)}{2}\mathbb{P}[Z \leq \underline{z}+s] + L(\phi)\mathbb{P}[Z > \underline{z}+s] \\
&= L(\phi) + \frac{1+l(\phi)-2L(\phi)}{2}\mathbb{P}[Z \leq \underline{z}+s] \\
&\leq L(\phi) + \frac{1+l(\phi)-2L(\phi)}{2}\mathbb{P}[Z \leq q_\alpha^-(Z)] \\
&\leq L(\phi) + \frac{1+l(\phi)-2L(\phi)}{2}\alpha,
\end{aligned} \tag{3.1.76}$$

using the fact that $-\phi'(z-s;1) \leq (1+l(\phi))/2$, for $z-s \leq \underline{z}$, and $-\phi'(z-s;1) \leq L(\phi)$ in the first inequality. And the fact that $\underline{z}+s \geq q_\alpha^-(Z)$ and $1-l(\phi) \leq 2L(\phi)$ in the second one. □

**Remark 3.1.16.** *In the previous lemma, we could have taken:*

$$\begin{aligned} \overline{z} &:= \inf\left\{z \in \mathbb{R} : \phi'(z;1) \geq \frac{1+L(\phi)}{2}\right\}, & (3.1.77) \\ \underline{z} &:= \sup\left\{z \in \mathbb{R} : -\phi'(z;-1) \leq \frac{1+l(\phi)}{2}\right\}. & (3.1.78) \end{aligned}$$

*Indeed, we just have to check that $\overline{z}$ and $\underline{z}$ belong, respectively, to the correspondents sets above. These facts follow from items (ii) and (iii) of Proposition 2.5.35, respectively.* □

Now we show that the solution set of the optimization problem associated with a general OCE risk measure satisfying some regularity conditions is a nonempty bounded interval.

**Proposition 3.1.17.** *Let be given $Z \in \mathcal{Z}$ and $\phi \in \Phi$ satisfying $l(\phi) < 1 < L(\phi) < +\infty$. The solution set of the convex optimization problem on the real numbers:*

$$\min_{s\in\mathbb{R}}\{\eta(s) := s + \mathbb{E}\phi(Z-s)\} \tag{3.1.79}$$

*is a nonempty closed interval that is contained in the interval:*

$$\left[q^-_{\gamma_1}(Z) - \overline{z}, q^+_{\gamma_2}(Z) - \underline{z}\right], \tag{3.1.80}$$

*where:*

$$\begin{aligned} \gamma_1 := \gamma_1(l(\phi), L(\phi)) &:= \frac{L(\phi)-1}{1+L(\phi)-2l(\phi)}, & (3.1.81) \\ \gamma_2 := \gamma_2(l(\phi), L(\phi)) &:= \frac{2(L(\phi)-1)}{2L(\phi)-1-l(\phi)}, & (3.1.82) \end{aligned}$$

*and $\overline{z}$ and $\underline{z}$ are defined as in (3.1.77) and (3.1.78), respectively.*

*Proof.* First of all, let us point out that $0 < \gamma_1 < \gamma_2 < 1$, since $0 \leq l(\phi) < 1 < L(\phi) < +\infty$. Since $\eta$ is a finite-valued convex function, it follows from Proposition 2.5.41 that:

$$\operatorname*{argmin}_{s\in\mathbb{R}} \eta(s) = \mathbb{R} \cap [\underline{s}, \overline{s}], \tag{3.1.83}$$

where $\underline{s} := \sup\{s \in \mathbb{R} : -\eta'(s;-1) < 0\}$ and $\overline{s} := \inf\{s \in \mathbb{R} : \eta'(s;1) > 0\}$[5]. Of course, this already shows that $\operatorname{argmin}_{s\in\mathbb{R}} \eta(s)$ is closed and convex[6].

Take $s \leq q^+_\alpha(Z) - \overline{z}$, where $0 < \alpha < \gamma_1$. We have that:

$$\begin{aligned} -\eta'(s;-1) &= 1 - \mathbb{E}\left[\phi'(Z-s;1)\right] \\ &\leq 1 - \frac{1+L(\phi)}{2}(1-\alpha) \\ &< 1 - l(\phi)\gamma_1 - \frac{1+L(\phi)}{2}(1-\gamma_1) = 0. \end{aligned} \tag{3.1.84}$$

[5]Until now, we have in principle that $\underline{s}, \overline{s} \in \overline{\mathbb{R}}$. We will show later that both $\underline{s}$ and $\overline{s}$ are finite.
[6]This is also an immediate consequence of the convexity and continuity of $\eta$.

We conclude that $q^+_\alpha(Z) - \overline{z} \leq \underline{s}$, for all $0 < \alpha < \gamma_1$. By item (iii) of Proposition 2.3.4 we have that:

$$q^-_{\gamma_1}(Z) = \sup_{\alpha<\gamma_1} q^+_\alpha(Z). \tag{3.1.85}$$

Therefore, $q^-_{\gamma_1}(Z) - \overline{z} \leq \underline{s}$.

Now, take $s \geq q^-_\alpha(Z) - \underline{z}$, where $\gamma_2 < \alpha < 1$ is arbitrary. We have that:

$$\begin{aligned} \eta'(s;1) &= 1 + \mathbb{E}\left[\phi'(Z-s;-1)\right] \\ &\geq 1 - L(\phi) - \frac{1 + l(\phi) - 2L(\phi)}{2}\alpha \\ &> 1 - L(\phi) - \frac{1 + l(\phi) - 2L(\phi)}{2}\gamma_2 = 0. \end{aligned} \tag{3.1.86}$$

We conclude that $\overline{s} \leq q^-_\alpha(Z) - \underline{z}$, for all $\gamma_2 < \alpha < 1$. By item (ii) of Proposition 2.3.4 we have that:

$$q^+_{\gamma_2}(Z) = \inf_{\alpha>\gamma_2} q^-_\alpha(Z). \tag{3.1.87}$$

Then, $\overline{s} \leq q^+_{\gamma_2}(Z) - \underline{z}$.

Summing up, we have proved that

$$\operatorname*{argmin}_{s\in\mathbb{R}} \eta(s) = [\underline{s}, \overline{s}] \subseteq \left[q^-_{\gamma_1}(Z) - \overline{z}, q^+_{\gamma_2}(Z) - \underline{z}\right]. \tag{3.1.88}$$

Now, we only have to show that $\operatorname{argmin}_{s\in\mathbb{R}} \eta(s) \neq \emptyset$, that is: $\underline{s} \leq \overline{s}$. This is a consequence of Proposition 2.5.35. In fact, let $s < \underline{s}$ be given. Taking $s < t < \underline{s}$, we obtain that:

$$\eta'(s;1) \leq \eta'(t;-1) < 0 \leq \lim_{s>\overline{s}} \eta'(s;1) = \eta'(\overline{s};1). \tag{3.1.89}$$

It follows that $s < \overline{s}$, i.e. $\underline{s} \leq \overline{s}$. □

In the next example we show that if the condition $L(\phi) < +\infty$ is not satisfied, then it is not possible, in general, to bound the set of optimal solutions of problem

$$\min_{s\in\mathbb{R}} \{s + \mathbb{E}\phi(Z-s)\} \tag{3.1.90}$$

using an interval whose extreme points depend on particular quantiles $0 < \gamma_1 < \gamma_2 < 1$ of $Z$ and on quantities $\underline{z}$ and $\overline{z}$ that depend only on $\phi$.

**Example 3.1.18.** *Take $\phi(z) := \exp(z) - 1$. It is elementary to verify that $\phi \in \Phi$. Moreover, since $\phi'(z) = \exp(z)$, we conclude that $l(\phi) = 0 < 1 < +\infty = L(\phi)$. Note that $\phi$ satisfy all conditions of Proposition 3.1.17, but the "$L(\phi) < +\infty$" condition. Take any $\alpha \in (0,1)$ and $\underline{z} \in \mathbb{R}$. We will show that there exists an integrable (in fact, bounded!) random variable $Z$, whose distribution depends on $\alpha$ and on $\underline{z}$, such that:*

$$\operatorname*{argmin}_{s\in\mathbb{R}} \{s + \mathbb{E}\phi(Z-s)\} \nsubseteq \left(-\infty, q^+_\alpha(Z) - \underline{z}\right]. \tag{3.1.91}$$

*Let us begin by showing that, for this particular $\phi$, the optimization problem in variable $s$ has as its unique solution $\bar{s} = \log(\mathbb{E}\exp(Z))$*[7]*. To see this, just set the derivative of (the differentiable convex) function*

$$s \in \mathbb{R} \mapsto s + \mathbb{E}\exp(Z - s) - 1 \tag{3.1.92}$$

*equal to zero and solve for $s$. Now, observe that:*

$$\bar{s} = \log(\mathbb{E}\exp(Z)) \geq \log(\exp \mathbb{E}Z) = \mathbb{E}Z, \tag{3.1.93}$$

*where we have used Jensen's inequality (see [22, Theorem 1.6.2]) above.*

*So, given $\alpha \in (0,1)$ and $\underline{z} \in \mathbb{R}$, we just need to define a random variable $Z$ such that:*

$$\mathbb{E}Z > q_\alpha^+(Z) - \underline{z}. \tag{3.1.94}$$

*Take $Z$ having the following probability distribution:*

$$\begin{aligned} \mathbb{P}[Z = 0] &= \frac{1+\alpha}{2}, & (3.1.95)\\ \mathbb{P}\left[Z = \frac{2|\underline{z}| + 2}{1-\alpha}\right] &= \frac{1-\alpha}{2}. & (3.1.96)\end{aligned}$$

*It is elementary to verify that:*

$$\begin{aligned} q_\alpha^+(Z) &= 0, \ \textit{and} & (3.1.97)\\ \mathbb{E}Z &= |\underline{z}| + 1. & (3.1.98)\end{aligned}$$

*The result follows, since:*

$$\mathbb{E}Z = |\underline{z}| + 1 > -\underline{z} = q_\alpha^+(Z) - \underline{z}. \tag{3.1.99}$$

□

The example below compares the estimate given by Proposition 3.1.17 of the set of optimal solutions of the optimization problem (3.1.79) with the exact set of optimal solutions for the $\mathrm{AV@R}_{1-\alpha}(\cdot)$ risk measure.

**Example 3.1.19.** *For $0 < \alpha < 1$, let us consider the $\mathrm{AV@R}_{1-\alpha}(\cdot)$ risk measure, i.e. take $\phi_\alpha(z) := \max\{0, z/(1-\alpha)\}$, for all $z \in \mathbb{R}$. As it is well-known (see also Proposition 3.1.13), for any $Z \in L_1(\Omega, \mathcal{F}, \mathbb{P})$, the set of $\alpha$-quantiles of $Z$ is the solution set of the convex optimization problem:*

$$\min_{s \in \mathbb{R}} \{s + \mathbb{E}[\phi_\alpha(Z - s)]\}. \tag{3.1.100}$$

[7] For the record, we also have that $\mu_\phi(Z) = \bar{s} + \mathbb{E}\phi(Z - \bar{s}) = \bar{s} = \log(\mathbb{E}\exp\{Z\})$, for every $Z \in \mathcal{Z}$.

*The left-side $\alpha$-quantile of $Z$ is the so-called Value-at-Risk risk measure. Observe that $l(\phi_\alpha) = 0 < 1 < 1/(1-\alpha) = L(\phi_\alpha) < +\infty$. Moreover, we can take $\underline{z} = 0 = \overline{z}$ such that the conditions below hold (see Lemma 3.1.15 and Remark 3.1.16):*

$$\phi'_\alpha(z;1) = \frac{1}{1-\alpha} \geq \frac{1-\alpha/2}{1-\alpha} = \frac{1+L(\phi_\alpha)}{2}, \text{ for all } z \geq \overline{z} = 0, \tag{3.1.101}$$

$$\phi'_\alpha(z;-1) = 0 \leq 1/2 = \frac{1+l(\phi_\alpha)}{2}, \text{ for all } z \leq \underline{z} = 0. \tag{3.1.102}$$

*So, by Proposition 3.1.17, we have that:*

$$\operatorname*{argmin}_{s\in\mathbb{R}}\{s + \mathbb{E}\phi_\alpha(Z-s)\} = \left[q_\alpha^-(Z), q_\alpha^+(Z)\right] \subseteq \left[q_{\gamma_1}^-(Z), q_{\gamma_2}^+(Z)\right], \tag{3.1.103}$$

*where:*

$$\gamma_1 = \frac{\alpha}{2-\alpha}, \tag{3.1.104}$$

$$\gamma_2 = \frac{2\alpha}{1+\alpha}. \tag{3.1.105}$$

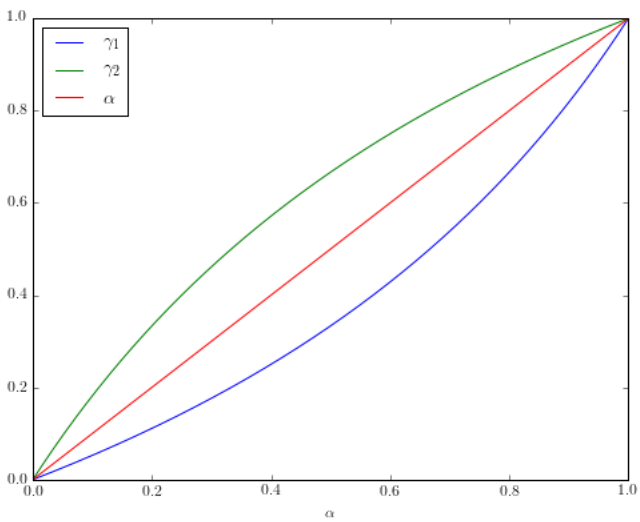


Figure 3.1: Lower and upper bounds for the solution set of problem (3.1.100) as quantiles of the random variable $Z$.

*Of course, $\gamma_1 \leq \alpha \leq \gamma_2$, for all $\alpha \in (0,1)$. Moreover, for $j = 1,2$, $\gamma_j \to 0$, as $\alpha \to 0$, and $\gamma_j \to 1$, as $\alpha \to 1$.* $\square$

## 3.2 Static problems with OCE risk measures

Let us consider a general static risk averse stochastic programming problem (SRA-SPP):

$$\min_{x \in X} \{v(x) := \mu_\phi(F(x,\xi))\} \tag{3.2.1}$$

with OCE risk measure $\mu_\phi$, with $\phi \in \Phi$. Akin to static risk neutral problems, $\xi = (\xi_1, \ldots, \xi_d)$ is a random vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$; $x \in \mathbb{R}^n$ are the decision variables; $X \subseteq \mathbb{R}^n$ is the feasible set and $F : \mathbb{R}^n \times \mathbb{R}^d \to \overline{\mathbb{R}}$ is a measurable function.

For risk neutral problems the expected value operator "$\mathbb{E}(\cdot)$" is used to summarize the random cost $F(x,\xi)$ into a real number $f(x) = \mathbb{E}F(x,\xi)$. This can be reasonable for many applications, although note that the random cost $F(x,\xi)$ can be much larger than its mean $f(x)$ for some realizations of $\xi$. In some situations it makes sense to give an extra penalization to upper deviations of $F(x,\xi)$ from its mean. Such problems use a risk measure to summarize the random cost into a real number.

For problems with OCE risk measures, a risk measure "$\mu_\phi(\cdot)$" is used to summarize the random cost $F(x,\xi)$ into a real number $v(x) = \mu_\phi(F(x,\xi))$. As we have seen in Example 3.1.8, the expected value operator is, in particular, an OCE risk measure. Moreover, given $\lambda \in [0,1]$ and $\alpha \in [0,1)$, we have that:

$$Z \in L_1(\Omega, \mathcal{F}, \mathbb{P}) \mapsto (1-\lambda)\mathbb{E}Z + \lambda\, \mathrm{AV@R}_{1-\alpha}(Z) \tag{3.2.2}$$

is an OCE risk measure. This is a mean-risk type of risk measure that is widely used in applications (e.g., [19, 39, 48, 51, 70, 72, 76, 78]).

For stochastic programming problems with OCE risk measures, one can also apply Monte Carlo methods in order to build a problem with discrete random data. We will continue to denote this approach as the SAA method. Note that the "average" is just the *sample* or *empirical* mean, so we will substitute the OCE risk measure by its *empirical* counterpart.

Given a sample realization $\{\xi^1, \ldots, \xi^N\}$ of $\xi$, one considers the empirical random vector $\hat{\xi}$ having the empirical distribution:

$$\hat{\mathbb{P}}\left[\hat{\xi} \in B\right] = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi^i}(B), \forall B \in \mathcal{B}(\mathbb{R}^d), \tag{3.2.3}$$

where:

$$\delta_y(B) := \begin{cases} 1, & \text{if } y \in B \\ 0, & \text{otherwise} \end{cases}.$$

Given $\phi \in \Phi$, the SAA problem is:

$$\min_{x \in X} \left\{\hat{v}_N(x) := \hat{\mu}_\phi\left(F(x,\hat{\xi})\right)\right\}, \tag{3.2.4}$$

where:

$$\hat{\mu}_\phi\left(F(x,\hat{\xi})\right) := \inf_{s\in\mathbb{R}}\left\{s+\hat{\mathbb{E}}\phi\left(F(x,\hat{\xi})-s\right)\right\} \tag{3.2.5}$$

$$= \inf_{s\in\mathbb{R}}\left\{s+\frac{1}{N}\sum_{i=1}^{N}\phi\left(F(x,\xi^i)-s\right)\right\} \tag{3.2.6}$$

Take $\phi \in \Phi$ and $x \in X$. Note that $\mu_\phi(F(x,\xi))$ is the optimal value of an optimization problem in $\mathbb{R}$. This motivates us to consider an extended formulation of problem (3.2.1):

$$\min_{(x,s)\in X\times\mathbb{R}}\left\{v(x,s) := s+\mathbb{E}\phi(F(x,\xi)-s)\right\}, \tag{3.2.7}$$

by adding an extra decision variable $s \in \mathbb{R}$. Given a random sample $\{\xi^1,\ldots,\xi^N\}$ of $\xi$, we can also consider an extended formulation of problem (3.2.4):

$$\min_{(x,s)\in X\times\mathbb{R}}\left\{\hat{v}_N(x,s) := s+\frac{1}{N}\sum_{i=1}^{N}\phi(F(x,\xi^i)-s)\right\}. \tag{3.2.8}$$

Note that[8]

$$v(x) = \inf_{s\in\mathbb{R}} v(x,s), \tag{3.2.9}$$

$$\hat{v}_N(x) = \inf_{s\in\mathbb{R}} \hat{v}_N(x,s). \tag{3.2.10}$$

One advantage of considering the extended formulation when studying the sample complexity of the SAA method is that the theory already developed for risk neutral problems can be applied. Indeed, observe that the expected value operator is the risk measure of the extended formulation. One difficult that should be circumvented is that the feasible set $X\times\mathbb{R}$ is unbounded, even when $X$ is bounded. The general theory developed for analyzing the sample complexity of risk neutral problems assumes a bounded feasible set (see assumption (A4) in section 2.1). Later we will see how it is possible to deal with the unboundedness of $X\times\mathbb{R}$, when $X$ is assumed bounded.

In the end of this subsection, we will show that $v(x,s)$, $v(x)$, $\hat{v}_N(x,s)$ and $\hat{v}_N(x)$ are all well-defined quantities, for all $x \in X$ and $s \in \mathbb{R}$, assuming that (A1) of subsection (2.1.1) holds. In the meantime, let us assume that all these quantities are well-defined[9].

[8]In our exposition we commit an abuse of notation by representing different functions $(x,s) \in X\times\mathbb{R} \mapsto s+\mathbb{E}\phi(F(x,\xi)-s)$ and $x \in X \mapsto \mu_\phi(F(x,\xi))$ with the same letter "$v$". It will be clear from the context which function we are talking about. The same remark applies to the objective-functions of the SAA problems: $(x,s) \in X\times\mathbb{R} \mapsto s+\sum_{i=1}^{N}\phi(F(x,\xi^i)-s)$ and $x \in X \mapsto \hat{\mu}_\phi(F(x,\hat{\xi}))$.

[9]Note that if $v(x,s)$ and $\hat{v}_N(x,s)$ are well-defined, for all $x \in X$ and $s \in \mathbb{R}$, then $v(x)$ and $\hat{v}_N(x)$ are well-defined, for all $x \in X$. The delicate point is that the expected value $\mathbb{E}\phi(F(x,\xi)-s)$ could fail, in principle, to be well-defined.

Let us denote the optimal values of problems (3.2.1) and (3.2.4), respectively, by:

$$v^* := \inf_{x \in X} v(x) \qquad \text{and} \qquad \hat{v}_N^* := \inf_{x \in X} \hat{v}_N(x). \tag{3.2.11}$$

It is elementary to verify that:

$$\begin{aligned} v^* &= \inf_{(x,s) \in X \times \mathbb{R}} v(x,s), & (3.2.12)\\ \hat{v}_N^* &= \inf_{(x,s) \in X \times \mathbb{R}} \hat{v}_N(x,s). & (3.2.13) \end{aligned}$$

Given $\epsilon \geq 0$, consider the following sets of optimal $\epsilon$-solutions of the problems above:

$$\begin{aligned} S^\epsilon &:= \{x \in X : v(x) \leq v^* + \epsilon\}, & (3.2.14)\\ ES^\epsilon &:= \{(x,s) \in X \times \mathbb{R} : v(x,s) \leq v^* + \epsilon\}, & (3.2.15)\\ \hat{S}_N^\epsilon &:= \{x \in X : \hat{v}_N(x) \leq \hat{v}_N^* + \epsilon\}, & (3.2.16)\\ \hat{ES}_N^\epsilon &:= \{(x,s) \in X \times \mathbb{R} : \hat{v}_N(x,s) \leq \hat{v}_N^* + \epsilon\}. & (3.2.17) \end{aligned}$$

In section 3.3 we will extend the sample complexity results obtained for static risk neutral problems to static problems with OCE risk measures under the same regularity conditions that were assumed for the risk neutral case, that is, assumptions (A1)-(A5) of subsection 2.1.1.

Similarly to risk neutral problems, we estimate the sample size $N$ such that:

$$\mathbb{P}\left(\left[\hat{S}_N^\delta \subseteq S^\epsilon\right] \cap \left[\hat{S}_N^\delta \neq \emptyset\right]\right) \geq 1 - \theta \tag{3.2.18}$$

holds true, where $0 \leq \delta < \epsilon$ and $\theta \in (0,1)$ are the sample complexity parameters. Additionally, we obtain an estimate for the sample size $N$ such that:

$$\mathbb{P}\left(\left[\hat{ES}_N^\delta \subseteq ES^\epsilon\right] \cap \left[\hat{ES}_N^\delta \neq \emptyset\right]\right) \geq 1 - \theta \tag{3.2.19}$$

also holds true.

Akin to the risk neutral case, we show that both the SAA and the extended SAA problems are solvable w.p.1, that is:

$$\mathbb{P}\left[\hat{S}_N \neq \emptyset\right] = \mathbb{P}\left[\hat{ES}_N \neq \emptyset\right] = 1. \tag{3.2.20}$$

For obtaining a lower bound estimate for the probability of the event:

$$\left[\hat{S}_N^\delta \subseteq S^\epsilon\right], \tag{3.2.21}$$

we show that the following probability:

$$\mathbb{P}\left[\sup_{x \in X} |\hat{v}_N(x) - v(x)| \leq \frac{\epsilon - \delta}{2}\right] \tag{3.2.22}$$

approaches 1 exponentially fast with respect to the sample size $N$. For risk neutral problems, this was shown by applying the uniform exponential bound theorem (see Theorem 2.1.5) that was used to bound from below:

$$\mathbb{P}\left[\sup_{x\in X}\left|\hat{f}_N(x)-f(x)\right|\leq\frac{\epsilon-\delta}{2}\right], \tag{3.2.23}$$

where $f(x)=\mathbb{E}F(x,\xi)$ and $\hat{f}_N(x)=\frac{1}{N}\sum_{i=1}^N F(x,\xi^i)$. Here, we cannot follow this approach directly, since the risk measure in problem (3.2.1) is not the expected value, in general. Indeed, [73, Proposition 5.6] and its previous discussion implies that:

$$\mathbb{E}\hat{v}_N(x)\leq v(x),\forall x\in X, \tag{3.2.24}$$

typically with strict inequality. One can try to show that $\hat{v}_N(x,s)$ and $v(x,s)$ become arbitrarily close, with high probability, uniformly on $X\times\mathbb{R}$. Since

$$v(x,s) = s+\mathbb{E}\phi(F(x,\xi)-s),\ \text{and} \tag{3.2.25}$$

$$\hat{v}_N(x,s) = s+\frac{1}{N}\sum_{i=1}^N\phi(F(x,\xi^i)-s) \tag{3.2.26}$$

the uniform exponential bound theorem could, in principle, be applied. The difficulty here is that $\operatorname{diam}(X\times\mathbb{R})=+\infty$ and the estimate from Theorem 2.1.5 is useless.

In order to deal with the unboundedness of the set $X\times\mathbb{R}$, we will introduce an auxiliary set $\tilde{X}\subseteq X\times\mathbb{R}$. Given $\tilde{X}$, we consider the extended formulations of the true and SAA problems restricted to this set. These new problems have the following optimal values:

$$v^*(\tilde{X}) := \inf_{(x,s)\in\tilde{X}} v(x,s), \tag{3.2.27}$$

$$\hat{v}_N^*(\tilde{X}) := \inf_{(x,s)\in\tilde{X}} \hat{v}_N(x,s). \tag{3.2.28}$$

Similarly, we denote the sets of $\epsilon$-solutions of the extended formulations restricted to $\tilde{X}$ by:

$$ES^\epsilon(\tilde{X}) := \left\{(x,s)\in X\times\mathbb{R}: v(x,s)\leq v^*(\tilde{X})+\epsilon\right\}, \tag{3.2.29}$$

$$\hat{ES}_N^\epsilon(\tilde{X}) := \left\{(x,s)\in X\times\mathbb{R}: \hat{v}_N(x,s)\leq\hat{v}_N^*(\tilde{X})+\epsilon\right\}, \tag{3.2.30}$$

respectively.

In Proposition 3.2.1 we show that if $\tilde{X}$ satisfies some suitable properties, then a series of useful consequences that relate problems (3.2.27) and (3.2.28), respectively, with problems (3.2.7) and (3.2.8) hold. Before proceeding, let us introduce some notation. We denote the projection of $\mathbb{R}^n\times\mathbb{R}$ on the first variable by:

$$\pi_x(x,s):=x,\ \text{for all}\ (x,s)\in\mathbb{R}^n\times\mathbb{R}. \tag{3.2.31}$$

Given $A \subseteq \mathbb{R}^n \times \mathbb{R}$ and $x \in \mathbb{R}^n$, we denote the cross-section of the set $A$ at the point $x$ by:

$$A_x := \{s \in \mathbb{R} : (x,s) \in A\}. \tag{3.2.32}$$

It is elementary to show that $a \in \pi_x(A)$ if and only if $A_a \neq \emptyset$.

**Proposition 3.2.1.** *Take $X \subseteq \mathbb{R}^n$ nonempty, a sample realization $\{\xi^1, \ldots, \xi^N\}$ of $\xi$ and $\tilde{X} \subseteq X \times \mathbb{R}$. Suppose that the following conditions hold:*

*(i) For every $x \in X$, $\tilde{X}_x \cap \operatorname{argmin}_{s\in\mathbb{R}} v(x,s) \neq \emptyset$.*

*(ii) For every $x \in X$, $\tilde{X}_x \cap \operatorname{argmin}_{s\in\mathbb{R}} \hat{v}_N(x,s) \neq \emptyset$.*

*Then, the following statements hold:*

*(a) The optimal values of problems (3.2.7) and (3.2.27) and of problems (3.2.8) and (3.2.28), respectively, are equal*

$$\begin{aligned} v^* &= v^*(\tilde{X}), \\ \hat{v}^*_N &= \hat{v}^*_N(\tilde{X}). \end{aligned} \tag{3.2.33}$$

*(b) For all $\epsilon \geq 0$, $S^\epsilon = \pi_x(ES^\epsilon(\tilde{X}))$ and $\hat{S}^\epsilon_N = \pi_x(\hat{ES}^\epsilon_N(\tilde{X}))$.*

*(c) Take $0 \leq \delta < \epsilon$. If $v^*$ is finite and*

$$\sup_{(x,s)\in\tilde{X}} |v(x,s) - \hat{v}_N(x,s)| \leq \frac{\epsilon - \delta}{2}, \tag{3.2.34}$$

*then $\hat{ES}^\delta_N(\tilde{X}) \subseteq ES^\epsilon(\tilde{X})$ and $\hat{S}^\delta_N \subseteq S^\epsilon$.*

*Proof.* Let us begin by noticing that $\pi_x(\tilde{X}) = X$ by item (i). Indeed, we have that $\tilde{X}_x \neq \emptyset$, for all $x \in X$. The fact that $v^* \leq v^*(\tilde{X})$ and $\hat{v}^*_N \leq \hat{v}^*_N(\tilde{X})$ is immediate, since $\tilde{X} \subseteq X \times \mathbb{R}$. Let us show the converse inequality. Take $x \in X$ arbitrary. By items (i) and (ii), there exist $s(x), \hat{s}(x) \in \tilde{X}_x$ such that:

$$v(x, s(x)) = v(x), \text{ and} \tag{3.2.35}$$

$$\hat{v}_N(x, \hat{s}(x)) = \hat{v}_N(x), \tag{3.2.36}$$

respectively. This implies that $v^*(\tilde{X}) \leq v(x)$ and $\hat{v}_N(\tilde{X}) \leq \hat{v}_N(x)$. Since $x \in X$ is arbitrary, we conclude that $v^*(\tilde{X}) \leq v^*$ and $\hat{v}^*_N(\tilde{X}) \leq \hat{v}^*_N$, which proves (a).

Let us show (b). Since the reasoning is similar for both cases, we will show that $S^\epsilon = \pi_x(ES^\epsilon(\tilde{X}))$, for any $\epsilon \geq 0$. Take any $(x,s) \in ES^\epsilon(\tilde{X})$. Observe that $v(x) \leq v(x,s) \leq v^*(\tilde{X}) + \epsilon = v^* + \epsilon$, since $v^*(\tilde{X}) = v^*$. It follows that $x \in S^\epsilon$, i.e. $\pi_x(ES^\epsilon(\tilde{X})) \subseteq S^\epsilon$. Conversely, take any $x \in S^\epsilon$. Taking $s(x) \in \tilde{X}_x \cap$

$\operatorname{argmin}_{s\in\mathbb{R}} v(x,s)$, we have that $v(x,s(x)) = v(x) \leq v^* + \epsilon = v^*(\tilde{X}) + \epsilon$, so $(x,s(x)) \in ES^{\epsilon}(\tilde{X})$, i.e. $x \in \pi_x(ES^{\epsilon}(\tilde{X}))$. Item (b) is proved.

Finally, take $0 \leq \delta < \epsilon$ and suppose that:

$$\sup_{(x,s)\in\tilde{X}} |v(x,s) - \hat{v}_N(x,s)| \leq \frac{\epsilon-\delta}{2} \tag{3.2.37}$$

holds. Since $\inf_{(x,s)\in\tilde{X}} v(x,s) = v^*$ is finite, it follows from Proposition 2.8.4 that:

$$\begin{aligned} \tfrac{\epsilon-\delta}{2} \geq \sup_{(x,s)\in\tilde{X}} |v(x,s) - \hat{v}_N(x,s)| &\geq \left| \inf_{(x,s)\in\tilde{X}} v(x,s) - \inf_{(x,s)\in\tilde{X}} \hat{v}_N(x,s) \right| \\ &= \left| v^*(\tilde{X}) - \hat{v}_N^*(\tilde{X}) \right| = |v^* - \hat{v}_N^*| \,. \end{aligned} \tag{3.2.38}$$

Given any $(x,s) \in \hat{ES}_N^{\delta}(\tilde{X})$, we have that:

$$v(x,s) - \frac{\epsilon-\delta}{2} \leq \hat{v}_N(x,s) \leq \hat{v}_N^* + \delta \leq v^* + \frac{\epsilon-\delta}{2} + \delta. \tag{3.2.39}$$

We conclude that $v(x,s) \leq v^* + \epsilon$, i.e. $(x,s) \in ES^{\epsilon}$. Then, $\hat{ES}_N^{\delta}(\tilde{X}) \subseteq ES^{\epsilon}(\tilde{X})$ and, in particular, $\hat{S}_N^{\delta} \subseteq S^{\epsilon}$ as a consequence of item (b). □

**Remark 3.2.2.** *Let us point out that items (i) and (ii) of Proposition 3.2.1 impose not only conditions to be satisfied by the set $\tilde{X}$, but also conditions regarding problems (3.2.7) and (3.2.8). Indeed, for every $x \in X$, the optimization problems on the real line:*

$$\min_{s\in\mathbb{R}} v(x,s), \quad \textit{and} \tag{3.2.40}$$

$$\min_{s\in\mathbb{R}} \hat{v}_N(x,s) \tag{3.2.41}$$

*are solvable. We will show in the next section that these conditions are satisfied under appropriate regularity conditions. Observe also that we do not assume in Proposition 3.2.1 that $\tilde{X}$ is bounded. We will show that it is possible to take $\tilde{X}$ bounded satisfying items (i) and (ii)[10] of that proposition. Then, we will apply the uniform bounded theorem to obtain an estimate on the sample size $N$ in order to bound from below the probability of the event:*

$$\sup_{(x,s)\in\tilde{X}} |v(x,s) - \hat{v}_N(x,s)| \leq \frac{\epsilon-\delta}{2}. \tag{3.2.42}$$

□

[10] More precisely, item (ii) will be satisfied for almost every random realization $\{\xi^1,\ldots,\xi^N\}$ of $\xi$.

**Remark 3.2.3.** *(Convention: Composition of Functions) Take two functions $g, h : \mathbb{R} \to \overline{\mathbb{R}}$. In general the composition $h \circ g$ is not well-defined, since $-\infty$ or $+\infty$ could be elements of the range of $g$. However we can extend the definition of $h$ to $\overline{\mathbb{R}}$ in the following way when both limits below do exist in $\overline{\mathbb{R}}$:*

$$h(-\infty) := \lim_{x \to -\infty} h(x), \tag{3.2.43}$$

$$h(+\infty) := \lim_{x \to +\infty} h(x). \tag{3.2.44}$$

*If $h$ is monotone, then both limits do exist. By item (c) of Definition 3.1.1 every $\phi \in \Phi$ is non-decreasing. Moreover, we have that $\phi(z) \geq z$, for all $z \in \mathbb{R}$. It follows that:*

$$\phi(-\infty) = \inf_{z \in \mathbb{R}} \phi(z), \tag{3.2.45}$$

$$\phi(+\infty) = \sup_{z \in \mathbb{R}} \phi(z) = +\infty. \tag{3.2.46}$$

*We will adopt this convention in this thesis.* □

**Remark 3.2.4.** *Take a random variable $Z$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Even if $Z$ has finite expected value, we do not rule out the possibility that $Z(\omega) = \pm\infty$, for some $\omega \in \Omega$. Nevertheless, a direct consequence from the finiteness of $\mathbb{E}Z$ is the fact that $\mathbb{P}[Z = \pm\infty] = 0$. By Remark 3.2.3 we have that $\phi(Z(\omega)) \in [\phi(-\infty), +\infty] \subseteq \overline{\mathbb{R}}$ is well-defined, for every $\omega \in \Omega$. Moreover, if $\phi$ is finite-valued, then $\mathbb{P}[\phi(Z) = \pm\infty] = 0$.* □

Let us recall that we have assumed until now that the objective functions of problems (3.2.1), (3.2.4), (3.2.7) and (3.2.8) are all well-defined. In the following proposition we give sufficient conditions for the well-definedness of these functions.

**Proposition 3.2.5.** *Suppose that each member of the family of random variables $\{F(x, \xi) : x \in X\}$ has finite expected value, i.e. assumption (A1) of Section 2.1.1 holds true. For any $\phi \in \Phi$, we have that:*

$$v(x, s) = s + \mathbb{E}\phi(F(x, \xi) - s), \text{ and} \tag{3.2.47}$$

$$v(x) = \mu_\phi(F(x, \xi)) \tag{3.2.48}$$

*are well-defined and belong to $\mathbb{R} \cup \{+\infty\}$, for every $x \in X$ and $s \in \mathbb{R}$. Moreover, if $L(\phi) < +\infty$, then both functions are finite on $X \times \mathbb{R}$ and on $X$, respectively. Finally, given any sample realization $\{\xi^1, \ldots, \xi^N\}$ of $\xi$, we have that:*

$$\hat{v}_N(x, s) := s + \frac{1}{N} \sum_{i=1}^{N} \phi\left(F(x, \xi^i) - s\right), \text{ and} \tag{3.2.49}$$

$$\hat{v}_N(x) := \hat{\mu}_\phi\left(F(x, \hat{\xi})\right) \tag{3.2.50}$$

*are well-defined functions.*

*Proof.* Take $\phi \in \Phi$ and a sample realization $\{\xi^1, \ldots, \xi^N\}$ of $\xi$. Let us begin by noting that, by the definition of the risk measures $\mu_\phi$ and $\hat{\mu}_\phi$, we have that:

$$v(x) = \inf_{s \in \mathbb{R}} v(x, s), \tag{3.2.51}$$

$$\hat{v}_N(x) = \inf_{s \in \mathbb{R}} \hat{v}_N(x, s), \tag{3.2.52}$$

for all $x \in X$. Given $x \in X$, $v(x)$ and $\hat{v}_N(x)$ will be well-defined whenever $v(x, s)$ and $\hat{v}_N(x, s)$ are well-defined, respectively, for all $s \in \mathbb{R}$. Since $F(x, \xi)$ has finite expected value, it follows from Proposition 3.1.4 that:

$$-\infty < \mathbb{E}F(x, \xi) \leq v(x) \leq s + \mathbb{E}\phi(F(x, \xi) - s) = v(x, s) \leq +\infty, \ \forall s \in \mathbb{R}. \tag{3.2.53}$$

Again by Proposition 3.1.4, we conclude that both functions are finite on $X \times \mathbb{R}$ and on $X$, respectively, if $\phi$ is Lispchitz continuous, i.e. $L(\phi) < +\infty$.

Now, let us consider an arbitrary sample realization $\{\xi^1, \ldots, \xi^N\}$ of $\xi$. Given $x \in X$, not much can be said about the scenario costs $\{F(x, \xi^i) : i = 1, \ldots, N\}$[11], but at least they satisfy:

$$-\infty \leq F(x, \xi^i) \leq +\infty, \ \forall i = 1, \ldots, N. \tag{3.2.54}$$

By Remark 3.2.3 it follows that $s + \phi(F(x, \xi^i) - s) \in \overline{\mathbb{R}}$, for all $s \in \mathbb{R}$ and $i = 1, \ldots, N$. Note that $\hat{v}_N(x, s)$ is just the mean of these quantities, that is well-defined, adopting the convention $+\infty + (-\infty) = +\infty$. $\square$

**Remark 3.2.6.** *In section 3.3, under additional regularity conditions, we show that (see Proposition 3.3.12), that:*

$$v : X \to \mathbb{R} \text{ and } v : X \times \mathbb{R} \to \mathbb{R} \tag{3.2.55}$$

*are Lipschitz continuous on $X$ and on $X \times \mathbb{R}$, respectively. Moreover, supposing that the random sample $\{\xi^i : i = 1, \ldots, N\}$ is identically distributed (not necessarily independent), we show in the same proposition that:*

$$\hat{v}_N : X \to \mathbb{R} \text{ and } \hat{v}_N : X \times \mathbb{R} \to \mathbb{R} \tag{3.2.56}$$

*are also Lipschitz continuous on $X$ and on $X \times \mathbb{R}$, respectively, w.p.1, where the Lipschitz constants depend on the sample realization $\{\xi^i : i = 1, \ldots, N\}$.* $\square$

[11] Assuming only that (A1) holds, it follows that there exists a measurable set $E_x$ such that $\mathbb{P}[\xi \in E_x] = 1$ and $F(x, \xi) \in \mathbb{R}$, for all $\xi \in E_x$. Since $X$ is, in general, an uncountable set, we need a stronger assumption in order to guarantee the finiteness of $F(x, \xi)$, for all $x \in X$ and $\xi \in E$, where $\mathbb{P}[\xi \in E_x] = 1$. See also Remark 2.1.1.

## 3.3 Sample complexity results for static problems

In this section, we apply the results developed in sections 3.1 and 3.2 for analyzing the sample complexity of static stochastic programming problems with OCE risk measures. We list below the same regularity conditions considered in subsection 2.1.1 which will be used in the sequel.

(A1) For every $x \in X$, $f(x) = \mathbb{E}F(x,\xi)$ is finite.

(A2) There exists $\sigma \in \mathbb{R}_+$ such that $F(x,\xi) - f(x)$ is a $\sigma$-sub-Gaussian random variable.

(A3) There exists a measurable function $\chi : \mathrm{supp}(\xi) \to \mathbb{R}_+$ whose moment generating function $M_\chi(s)$ is finite, for $s$ in a neighborhood of zero, such that

$$|F(x,\xi) - F(x',\xi)| \le \chi(\xi)\,\|x - x'\|\,, \tag{3.3.1}$$

for all $x', x \in X$ and $\xi \in E \subseteq \mathrm{supp}\{\xi\}$, where $\mathbb{P}\left[\xi \in E\right] = 1$.

(A4) $X \subseteq \mathbb{R}^n$ is a nonempty compact set with diameter $D$.

(A5) $\{\xi^i : i \in \mathbb{N}\}$ is an i.i.d. sequence of random vectors defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\xi^1 \overset{d}{\sim} \xi$.

As usual, given a sample realization $\{\xi^1, \ldots, \xi^N\}$ of $\xi$, we denote by $\hat{\xi}$ the random vector having the empirical distribution (3.2.3). We begin by proving the following lemma.

**Lemma 3.3.1.** *Take $\phi \in \Phi$, and a sample realization $\{\xi^1, \ldots, \xi^N\}$ of $\xi$ and $x \in X$. If $F(x,\xi^i) \in \mathbb{R}$, for every $1 \le i \le N$, then $\hat{v}_N(x,\cdot)$ is a well-defined proper convex function taking values in $\mathbb{R} \cup \{+\infty\}$. If $\phi$ is Lipschitz continuous, then $\hat{v}_N(x,\cdot)$ is finite-valued. Finally, if conditions $l(\phi) < 1 < L(\phi)$ also hold true, then*

$$\operatorname*{argmin}_{s \in \mathbb{R}} \hat{v}_N(x,s) \subseteq \left[q^-_{\gamma_1}\left(F(x,\hat{\xi})\right) - \overline{z},\, q^+_{\gamma_2}\left(F(x,\hat{\xi})\right) - \underline{z}\right] \tag{3.3.2}$$

*is a nonempty compact interval of $\mathbb{R}$, where $\gamma_1$ and $\gamma_2$ are defined as in equations* (3.1.81) *and* (3.1.82)*, respectively, and $\underline{z} \le 0 \le \overline{z}$ are constants defined as in equations* (3.1.77) *and* (3.1.78)*, respectively.*

*Proof.* Note that $\phi \in \Phi$ is a proper convex function satisfying $\mathrm{dom}\,\phi \supseteq (-\infty, 0]$. Moreover, $\phi(z) \in \mathbb{R} \cup \{+\infty\}$, for every $z \in \mathbb{R}$. Since $F(x,\xi^i) \in \mathbb{R}$, for every $1 \le i \le N$, we have that

$$s \in \mathbb{R} \mapsto s + \phi(F(x,\xi^i) - s) \tag{3.3.3}$$

is a proper convex function taking values in $\mathbb{R} \cup \{+\infty\}$ and whose domain contains the interval $[F(x,\xi^i), +\infty)$. Since $\hat{v}_N(x,\cdot)$ is a convex combination of the $N$ functions above, we conclude that $\hat{v}_N(x,\cdot)$ is convex, takes values in $\mathbb{R} \cup \{+\infty\}$ and satisfies:

$$\operatorname{dom} \hat{v}_N(x,\cdot) \supseteq \bigcap_{i=1}^{N} \left[F(x,\xi^i), +\infty\right) = \left[\max_{1\leq i\leq N} F(x,\xi^i), +\infty\right). \tag{3.3.4}$$

In particular, it follows that $\hat{v}_N(x,\cdot)$ is proper. Now, suppose that $\phi$ is Lipschitz continuous. It follows that $\phi$ is finite-valued and also each one of the functions in equation (3.3.3). So, $\hat{v}_N(x,\cdot)$ is also finite-valued.

Now, let us suppose that conditions $l(\phi) < 1 < L(\phi) < +\infty$ hold true. Let $\hat{\xi}$ be the empirical random vector having the empirical distribution (3.2.3). Note that the random variable $F(x,\hat{\xi})$ has finite expected valued $\hat{\mathbb{E}}F(x,\hat{\xi}) = \frac{1}{N}\sum_{i=1}^{N} F(x,\xi^i)$, since $F(x,\xi^i) \in \mathbb{R}$ ($\forall i = 1,\ldots,N$). By Proposition 3.1.17, we conclude that the solution set of problem:

$$\min_{s\in\mathbb{R}} \left\{ s + \hat{\mathbb{E}}\phi(F(x,\hat{\xi}) - s) \right\} \tag{3.3.5}$$

is a nonempty compact interval of $\mathbb{R}$ that is contained in

$$\left[q^-_{\gamma_1}\left(F(x,\hat{\xi})\right) - \overline{z}, q^+_{\gamma_2}\left(F(x,\hat{\xi})\right) - \underline{z}\right].$$

The result follows since $\hat{v}_N(x,s) = s + \hat{\mathbb{E}}\phi(F(x,\hat{\xi}) - s)$. □

In the sequel, we will define an auxiliary set $\tilde{X} \subseteq X \times \mathbb{R}$ having "good properties" that will allow us to derive sample complexity estimates for stochastic programming problems with OCE risk measures. Meanwhile, let us write:

$$\tilde{X} := \{(x,s) \in X \times \mathbb{R} : a(x) \leq s \leq b(x)\}, \tag{3.3.6}$$

where $a : X \to \mathbb{R}$ and $b : X \to \mathbb{R}$ will be defined along the way. The following proposition will be useful.

**Proposition 3.3.2.** *Assume that (A1) holds true and that $\phi \in \Phi$ satisfies $l(\phi) < 1 < L(\phi) < +\infty$. Given $x \in X$ and real numbers $a(x) < b(x)$, define:*

$$\Delta(x) := \frac{1}{3}\min\left\{v(x,a(x)) - v(x), v(x,b(x)) - v(x)\right\}. \tag{3.3.7}$$

*If $(a(x),b(x)) \supseteq \operatorname{argmin}_{s\in\mathbb{R}} v(x,s)$, then $\Delta(x) > 0$. Additionally, let be given a sample realization $\{\xi^1,\ldots,\xi^N\}$ of $\xi$ such that $F(x,\xi^i) \in \mathbb{R}$, for all $i = 1,\ldots,N$. Suppose also that there exists $\bar{s} \in \operatorname{argmin}_{s\in\mathbb{R}} v(x,s)$ such that:*

$$\max\left\{\left|\hat{v}_N(x,a(x)) - v(x,a(x))\right|, \left|\hat{v}_N(x,b(x)) - v(x,b(x))\right|, \left|\hat{v}_N(x,\bar{s}) - v(x)\right|\right\} < \Delta(x). \tag{3.3.8}$$

*Then,*

$$\left\{s \in \mathbb{R} : \hat{v}_N(x,s) \leq \hat{v}_N(x) + \Delta(x)\right\} \subseteq (a(x),b(x)).$$

*Proof.* Since $\phi \in \Phi$ satisfies $l(\phi) < 1 < L(\phi) < +\infty$ and the random variable $F(x,\xi)$ has finite expected value, it follows from Proposition 3.1.17 that $\operatorname{argmin}_{s\in\mathbb{R}} v(x,s)$ is a nonempty compact interval of $\mathbb{R}$. By hypothesis, $(a(x), b(x)) \supseteq \operatorname{argmin}_{s\in\mathbb{R}} v(x,s)$, so $a(x), b(x) \notin \operatorname{argmin}_{s\in\mathbb{R}} v(x,s)$. This implies that $\Delta(x) > 0$.

Now, let $\{\xi^1, \ldots, \xi^N\}$ be a sample realization of $\xi$ satisfying $F(x,\xi^i) \in \mathbb{R}$, for all $i = 1, \ldots, N$. By Lemma 3.3.1, it follows that $\hat{v}_N(x,\cdot)$ is a finite-valued convex function. If there exists $\bar{s} \in \operatorname{argmin}_{s\in\mathbb{R}} v(x,s)$ such that equation (3.3.8) is satisfied, then:

$$
\begin{aligned}
\hat{v}_N(x, a(x)) &> v(x, a(x)) - \Delta(x) && (3.3.9)\\
&\geq v(x, a(x)) - \frac{1}{3}\left(v(x, a(x)) - v(x)\right) && (3.3.10)\\
&= v(x) + \frac{2}{3}\left(v(x, a(x)) - v(x)\right) && (3.3.11)\\
&\geq v(x) + 2\Delta(x) = (v(x,\bar{s}) + \Delta(x)) + \Delta(x) && (3.3.12)\\
&> \hat{v}_N(x,\bar{s}) + \Delta(x) && (3.3.13)\\
&\geq \hat{v}_N(x) + \Delta(x). && (3.3.14)
\end{aligned}
$$

Similarly, we can show that $\hat{v}_N(x, b(x)) > \hat{v}_N(x) + \Delta(x)$. Now, observe that any $s < a(x)$ can be written as $\lambda a(x) + (1-\lambda)\bar{s}$, for some $\lambda > 1$. Since $\hat{v}_N(x,\cdot)$ is convex, it follows from Proposition 2.5.44 that:

$$
\begin{aligned}
\hat{v}_N(x,s) &\geq \lambda \hat{v}_N(x, a(x)) + (1-\lambda)\hat{v}_N(x,\bar{s}) && (3.3.15)\\
&= \hat{v}_N(x, a(x)) + (\lambda - 1)(\hat{v}_N(x, a(x)) - \hat{v}_N(x,\bar{s})) && (3.3.16)\\
&> \hat{v}_N(x, a(x)) && (3.3.17)\\
&> \hat{v}_N(x) + \Delta(x). && (3.3.18)
\end{aligned}
$$

Similarly, observe that any $s > b(x)$ can be written as $\lambda b(x) + (1-\lambda)\bar{s}$, for some $\lambda > 1$. The same reasoning implies that

$$
\hat{v}_N(x,s) \;>\; \hat{v}_N(x, b(x)) > \hat{v}_N(x) + \Delta(x), \tag{3.3.19}
$$

for $s > b(x)$. We conclude that $\{s \in \mathbb{R} : \hat{v}_N(x,s) \leq \hat{v}_N(x) + \Delta(x)\} \subseteq (a(x), b(x))$. □

**Remark 3.3.3.** *We will define $a, b : X \to \mathbb{R}$ such that the following conditions hold:*

*(i)* $\operatorname{int} \tilde{X}_x = (a(x), b(x)) \supseteq \operatorname*{argmin}_{s\in\mathbb{R}} v(x,s)$*, for all* $x \in X$*,*

*(ii)* $-\infty < \inf_{x\in X} a(x) < \sup_{x\in X} b(x) < +\infty$*,*

*(iii)* $\Delta := \inf_{x\in X} \Delta(x) > 0$*, where* $\Delta(x)$ *is defined as in equation (3.3.7).*

□

Observe that if $F(x,\xi)$ has finite expected value, for all $x \in X$, and $\phi \in \Phi$ satisfies $l(\phi) < 1 < L(\phi) < +\infty$, then assuming that $a(x)$ and $b(x)$ satisfy

$$a(x) < q^-_{\gamma_1}(F(x,\xi)) - \overline{z} \leq q^+_{\gamma_2}(F(x,\xi)) - \underline{z} < b(x), \tag{3.3.20}$$

we obtain by Proposition 3.1.17 that $(a(x), b(x)) \supseteq \operatorname*{argmin}_{s\in\mathbb{R}} v(x,s)$, for all $x \in X$. Of course, $\gamma_1$, $\gamma_2$, $\underline{z}$ and $\overline{z}$ depend only on $\phi$, see Proposition 3.1.17. This addresses item $(i)$ above. In the next proposition, we address item (iii). Later it will become clear that our sample complexity estimates hold for $\epsilon > 0$ sufficiently small, that is, less than or equal to $\Delta$. So, it is crucial to show that we can define $a(x)$ and $b(x)$ in such a way that $\Delta > 0$ and $\tilde{X}$ is bounded.

In the sequel, $\gamma_1$, $\gamma_2$, $\underline{z}$ and $\overline{z}$ are defined as in Proposition 3.1.17, where $\phi \in \Phi$. As before, we will write $f(x) := \mathbb{E}F(x,\xi)$ whenever the expected value is well-defined.

**Proposition 3.3.4.** *Assume that (A1) holds true and that there exist two functions $\mathfrak{l}, \mathfrak{u} : (0,1) \to \mathbb{R}$ satisfying the following conditions:*

*(i.) For all $x \in X$ and $\alpha \in (0,1)$,*

$$f(x) + \mathfrak{l}(\alpha) \leq q^-_\alpha(F(x,\xi)) \leq q^+_\alpha(F(x,\xi)) \leq f(x) + \mathfrak{u}(\alpha).$$

*(ii.) $\mathfrak{l}(\cdot)$ and $\mathfrak{u}(\cdot)$ are monotonically (strictly) increasing.*

*Let $\phi \in \Phi$ be such that $l(\phi) < 1 < L(\phi) < +\infty$ is satisfied. For every $x \in X$, define:*

$$a(x) := f(x) + \mathfrak{l}\left(\frac{\gamma_1}{2}\right) - \overline{z}, \tag{3.3.21}$$

$$b(x) := f(x) + \mathfrak{u}\left(\frac{1+\gamma_2}{2}\right) - \underline{z}, \tag{3.3.22}$$

*and $\Delta(x)$ as in equation (3.3.7). Then,*

$$\Delta := \inf_{x\in X} \Delta(x) \geq \tfrac{1}{3}\min\left\{\tfrac{L(\phi)-1}{8}\left[\mathfrak{l}(\tfrac{3\gamma_1}{4}) - \mathfrak{l}(\tfrac{\gamma_1}{2}))\right], \tfrac{1-l(\phi)}{8}\left[\mathfrak{u}(\tfrac{1+\gamma_2}{2}) - \mathfrak{u}(\tfrac{1}{4} + \tfrac{3\gamma_2}{4})\right]\right\} > 0. \tag{3.3.23}$$

*Proof.* Take any $x \in X$. Since $\phi$ is Lipschitz continuous and $F(x,\xi)$ has finite expected value, we conclude that $v(x,\cdot)$ is a finite-valued convex function in $\mathbb{R}$. So, for every $a < b$, Theorem 2.5.40 implies that:

$$v(x,b) - v(x,a) = \int_a^b v'_s(x,s)ds = \int_a^b v'_s(x,s;1)ds = \int_a^b -v'_s(x,s;-1)ds, \tag{3.3.24}$$

where $v'_s(x,s)$ is well-defined, except possibly in a countable subset of $[a,b]$. Moreover, by Proposition 3.1.17, we have that:

$$\emptyset \neq \operatorname*{argmin}_{s\in\mathbb{R}} v(x,s) \subseteq \left[q^-_{\gamma_1}(F(x,\xi) - \overline{z}, q^+_{\gamma_2}(F(x,\xi) - \underline{z}\right]. \tag{3.3.25}$$

Let $\bar{s} \in \operatorname{argmin}_{s\in\mathbb{R}} v(x,t)$ be arbitrary. Then,

$$\begin{aligned}
v(x,b(x)) - v(x) &= v(x,b(x)) - v(x,\bar{s}) && (3.3.26)\\
&= \int_{\bar{s}}^{b(x)} v_s'(x,s;1)ds && (3.3.27)\\
&\geq \int_{f(x)+\mathfrak{u}\left(\frac{1+3\gamma_2}{4}\right)-\underline{z}}^{f(x)+\mathfrak{u}\left(\frac{1+\gamma_2}{2}\right)-\underline{z}} v_s'(x,s;1)ds && (3.3.28)\\
&\geq \begin{array}{l}\left[\mathfrak{u}\left(\frac{1+\gamma_2}{2}\right) - \mathfrak{u}\left(\frac{1+3\gamma_2}{4}\right)\right] \times \\ \qquad v_s'\left(x, f(x) + \mathfrak{u}\left(\frac{1+3\gamma_2}{4}\right) - \underline{z}; 1\right).\end{array} && (3.3.29)
\end{aligned}$$

Observe that (3.3.28) holds since $\bar{s} \leq q^+_{\gamma_2}(F(x,\xi)) - \underline{z} \leq f(x) + \mathfrak{u}(\gamma_2) - \underline{z} \leq f(x) + \mathfrak{u}\left(\frac{1+3\gamma_2}{4}\right) - \underline{z}$ and $v_s'(x,s;1) \geq 0$, for all $s \geq \bar{s}$. Moreover, (3.3.29) holds since the right derivative $v_s'(x,\cdot;1)$ is non-decreasing by the convexity of $v(x,\cdot)$. Furthermore, since $\tilde{s} := f(x)+\mathfrak{u}\left(\frac{1+3\gamma_2}{4}\right)-\underline{z} \geq q^-_{\frac{1+3\gamma_2}{4}}(F(x,\xi))-\underline{z}$, it follows by item (iii) of Lemma 3.1.15 that

$$\begin{aligned}
v_s'\left(x,\tilde{s};1\right) &= 1 + \mathbb{E}\left[\phi'\left(F(x,\xi) - \tilde{s}; -1\right)\right] && (3.3.30)\\
&\geq 1 - L(\phi) - \frac{1 + l(\phi) - 2L(\phi)}{2}\left(\frac{1+3\gamma_2}{4}\right) && (3.3.31)\\
&= A_1\left(\frac{1+3\gamma_2}{4}\right), && (3.3.32)
\end{aligned}$$

where $A_1(\alpha) := 1 - L(\phi) - \frac{1+l(\phi)-2L(\phi)}{2}\alpha$. Observe that $A(\cdot)$ is an affine function satisfying $A(\gamma_2) = 0$, so

$$\begin{aligned}
A_1\left(\frac{1+3\gamma_2}{4}\right) &= \frac{1}{4}A_1(1) + \frac{3}{4}A_1(\gamma_2) && (3.3.33)\\
&= \frac{A_1(1)}{4} = \frac{1-l(\phi)}{8}. && (3.3.34)
\end{aligned}$$

We conclude that

$$v(x,b(x)) - v(x) \geq \frac{1-l(\phi)}{8}\left[\mathfrak{u}\left(\frac{1+\gamma_2}{2}\right) - \mathfrak{u}\left(\frac{1+3\gamma_2}{4}\right)\right] > 0, \qquad (3.3.35)$$

for every $x \in X$.

Similarly, taking any $x \in X$, let us obtain a lower estimate for $v(x, a(x)) - v(x)$:

$$
\begin{aligned}
v(x,a(x)) - v(x) &= v(x,a(x)) - v(x,\bar{s}) && (3.3.36)\\
&= \int_{\bar{s}}^{a(x)} -v_s'(x,s;-1)ds && (3.3.37)\\
&= \int_{a(x)}^{\bar{s}} v_s'(x,s;-1)ds && (3.3.38)\\
&\geq \int_{f(x)+\mathfrak{l}\left(\frac{\gamma_1}{2}\right)-\bar{z}}^{f(x)+\mathfrak{l}\left(\frac{3\gamma_1}{4}\right)-\bar{z}} v_s'(x,s;-1)ds && (3.3.39)\\
&\geq \left[\mathfrak{l}\left(\tfrac{3\gamma_1}{4}\right) - \mathfrak{l}\left(\tfrac{\gamma_1}{2}\right)\right] \times \\
&\qquad v_s'\left(x, f(x) + \mathfrak{l}\left(\tfrac{3\gamma_1}{4}\right) - \bar{z}; -1\right). && (3.3.40)
\end{aligned}
$$

Observe that (3.3.39) holds since $\bar{s} \geq q_{\gamma_1}^{-}(F(x,\xi)) - \bar{z} \geq f(x) + \mathfrak{u}(\gamma_1) - \bar{z} \geq f(x) + \mathfrak{u}\left(\frac{3\gamma_1}{4}\right) - \bar{z}$ and $v_s'(x,s;-1) \geq 0$, for all $s \leq \bar{s}$. Moreover, (3.3.40) holds since derivative $v_s'(x,\cdot;-1)$ is non-increasing by the convexity of $v(x,\cdot)$. Furthermore, since $\hat{s} := f(x) + \mathfrak{l}\left(\frac{3\gamma_1}{4}\right) - \bar{z} \leq q_{\frac{3\gamma_1}{4}}^{+}(F(x,\xi)) - \bar{z}$, it follows by item (ii) of Lemma 3.1.15 that

$$
\begin{aligned}
v_s'(x,\hat{s};-1) &= -1 + \mathbb{E}\left[\phi'\left(F(x,\xi) - \hat{s};1\right)\right] && (3.3.41)\\
&\geq -1 + \frac{1+L(\phi)}{2}\left(1 - \frac{3\gamma_1}{4}\right) && (3.3.42)\\
&= A_2\left(\frac{3\gamma_1}{4}\right), && (3.3.43)
\end{aligned}
$$

where $A_2(\alpha) := -1 + \frac{1+L(\phi)}{2}(1-\alpha)$. Observe that $A_2(\cdot)$ is an affine function and $A(\gamma_1) = 0$, so

$$
\begin{aligned}
A_2\left(\frac{3\gamma_1}{4}\right) &= \frac{1}{4}A_2(0) + \frac{3}{4}A_2(\gamma_1) && (3.3.44)\\
&= \frac{A_2(0)}{4} = \frac{L(\phi)-1}{8}. && (3.3.45)
\end{aligned}
$$

We have that

$$
v(x,a(x)) - v(x) \geq \frac{L(\phi)-1}{8}\left[\mathfrak{l}\left(\frac{3\gamma_1}{4}\right) - \mathfrak{l}\left(\frac{\gamma_1}{2}\right)\right] > 0, \tag{3.3.46}
$$

for every $x \in X$, which proves equation (3.3.23). $\square$

Observe that if: (a) $f : X \to \mathbb{R}$ is bounded, (b) conditions (i) and (ii) of Proposition 3.3.4 hold, and (c) we define $a(x)$ and $b(x)$ as in equations (3.3.21) and (3.3.22), respectively, then item (ii) of Remark 3.3.3 is satisfied. Let us recall that assumptions (A1), (A3) and (A4) imply the boundedness of $f$ on $X$ (see Proposition

2.1.4). Moreover, we will show that assumption (A2) guarantees the existence of functions $\mathfrak{l}, \mathfrak{u} : (0,1) \to \mathbb{R}$ satisfying items (i) and (ii) of Proposition 3.3.4.

In the next proposition we establish lower and upper bounds, respectively, for the leftmost and rightmost $\alpha$-quantile of a $\sigma$-sub-Gaussian random variable, for $\alpha \in (0,1)$.

**Proposition 3.3.5.** *Let $Y$ be a $\sigma$-sub-Gaussian random variable. For $\alpha \in (0,1)$, we have that*

$$-\sigma\sqrt{2\log\left(\frac{1}{\alpha}\right)} \le q_\alpha^-(Y) \le q_\alpha^+(Y) \le \sigma\sqrt{2\log\left(\frac{1}{1-\alpha}\right)}. \tag{3.3.47}$$

*Proof.* Let $y > 0$ be given. For $s > 0$ arbitrary, we have that

$$\mathbb{P}[Y \ge y] = \mathbb{P}[\exp\{sY\} \ge \exp\{sy\}] \le \mathbb{E}[\exp\{sY\}]\exp\{-sy\} \le \exp\{-(sy - \sigma^2 s^2/2)\}, \tag{3.3.48}$$

where the first inequality is just Markov inequality (see [22, Theorem 1.6.4]). Minimizing this expression with respect to $s > 0$, we obtain that

$$\mathbb{P}[Y \ge y] \le \exp\{-y^2/2\sigma^2\} = 1 - (1 - \exp\{-y^2/2\sigma^2\}). \tag{3.3.49}$$

Given $\alpha \in (0,1)$, we can take $y > 0$ such that

$$\alpha < 1 - \exp\{-y^2/2\sigma^2\}. \tag{3.3.50}$$

In fact, just take

$$y > \sigma\sqrt{2\log\left(\frac{1}{1-\alpha}\right)}. \tag{3.3.51}$$

For such $y$, we have that $\mathbb{P}[Y \ge y] < 1 - \alpha$. Since the function $z \mapsto \mathbb{P}[Y \ge z]$ is non-increasing and $\mathbb{P}[Y \ge q_\alpha^+(Y)] \ge 1 - \alpha$, we conclude that $q_\alpha^+(Y) < y$, for all $y$ satisfying (3.3.51). Taking the infimum on $y$, we obtain that

$$q_\alpha^+(Y) \le \sigma\sqrt{2\log\left(\frac{1}{1-\alpha}\right)}, \tag{3.3.52}$$

and the rightmost inequality on (3.3.47) is proved. The leftmost inequality is a direct consequence of this result, noting that $-Y$ is also an $\sigma$-sub-Gaussian random variable and that $q_\alpha^-(Y) = -q_{1-\alpha}^+(-Y)$ by Proposition 2.3.5. □

**Corollary 3.3.6.** *If assumptions (A1) and (A2) hold true, then*

$$f(x) - \sigma\sqrt{2\log\left(\frac{1}{\alpha}\right)} \le q_\alpha^-(F(x,\xi)) \le q_\alpha^+(F(x,\xi)) \le f(x) + \sigma\sqrt{2\log\left(\frac{1}{1-\alpha}\right)}, \tag{3.3.53}$$

*for every $\alpha \in (0,1)$ and $x \in X$.*

*Proof.* Let $\alpha \in (0,1)$ and $x \in X$ be arbitrary. The result follows immediately from Proposition 3.3.5 by taking $Y = F(x,\xi) - f(x)$ and observing that $q^-_\alpha(F(x,\xi) - f(x)) = q^-_\alpha(F(x,\xi)) - f(x)$ and $q^+_\alpha(F(x,\xi) - f(x)) = q^+_\alpha(F(x,\xi)) - f(x)$ hold. $\square$

It follows from assumptions (A1) and (A2) that if we take

$$\mathfrak{l}(\alpha) \;:=\; -\sigma\sqrt{2\log\left(\frac{1}{\alpha}\right)}, \text{ and} \tag{3.3.54}$$

$$\mathfrak{u}(\alpha) \;:=\; \sigma\sqrt{2\log\left(\frac{1}{1-\alpha}\right)} \tag{3.3.55}$$

then items (i) and (ii) of Proposition 3.3.4 are satisfied. Let us summarize this result in the next corollary.

**Corollary 3.3.7.** *Suppose that assumptions (A1) and (A2) are satisfied and that $\phi \in \Phi$ is such that $l(\phi) < 1 < L(\phi) < +\infty$. Defining $\mathfrak{l}(\cdot)$ and $\mathfrak{u}(\cdot)$ as in equations (3.3.54) and (3.3.55), respectively; and $a(x)$, $b(x)$ and $\Delta(x)$ as in Proposition 3.3.4, we conclude that $\Delta := \inf_{x\in X}\Delta(x) > 0$.*

*Proof.* This is an elementary consequence of Proposition 3.3.4 and Corollary 3.3.6. $\square$

Now, let us prepare the ground to show that, under appropriate regularity conditions, the probability that the objective functions

$$v(x,s) \;=\; s + \mathbb{E}\phi(F(x,\xi) - s), \text{ and} \tag{3.3.56}$$

$$\hat{v}_N(x,s) \;=\; s + \frac{1}{N}\sum_{i=1}^{N}\phi(F(x,\xi^i) - s) \tag{3.3.57}$$

become arbitrarily close on $\tilde{X}$ approaches one exponentially fast with respect to the sample size $N$. First of all, note that:

$$\hat{v}_N(x,s) - v(x,s) = \frac{1}{N}\sum_{i=1}^{N}\left[\phi(F(x,\xi^i) - s) - \mathbb{E}\phi(F(x,\xi^i) - s)\right], \tag{3.3.58}$$

assuming that $\xi^i \overset{d}{\sim} \xi$, for every $i = 1,\ldots,N$. So, in order for applying the uniform exponential bound theorem, we need to verify that the family of random variables $\{\phi(F(x,\xi)-s) : x \in X, s \in \mathbb{R}\}$ satisfies the regularity conditions of Theorem 2.1.5[12]. If $\phi \in \Phi$ satisfies $L(\phi) < +\infty$ and assumptions (A1)-(A3) hold true, then we will show that this will be the case.

[12]In fact, we only need to verify that those conditions are satisfied for the subfamily $\{\phi(F(x,\xi) - s) : (x,s) \in \tilde{X}\}$. It takes the same work to show the result for the whole family of random variables, though.

**Proposition 3.3.8.** *Take any $\phi \in \Phi$ satisfying $L(\phi) < +\infty$. If assumptions $(A1)-(A3)$ hold true, then the following conditions are satisfied:*

*(C1)* $\mathbb{E}\phi(F(x,\xi)-s)$ *is finite, for all $x \in X$ and $s \in \mathbb{R}$.*

*(C2) The family of random variables*

$$\{\phi(F(x,\xi)-s)-\mathbb{E}\phi(F(x,\xi)-s) : x \in X, s \in \mathbb{R}\}$$

*is $(\kappa L(\phi)\sigma)$-sub-Gaussian, where $\kappa$ is an absolute constant*[13].

*(C3) Defining*

$$\eta(\xi) := \sqrt{1+\chi(\xi)^2} \tag{3.3.59}$$

*we obtain that*

$$|\phi\left(F(x',\xi)-s'\right)-\phi\left(F(x,\xi)-s\right)| \leq L(\phi)\eta(\xi)\left\|(x',s')-(x,s)\right\|, \tag{3.3.60}$$

*for all $(x,s),(x',s') \in X \times \mathbb{R}$ and $\xi \in E \subseteq \mathrm{supp}\{\xi\}$, where $\mathbb{P}[\xi \in E] = 1$. Moreover, the moment generating function of $L(\phi)\eta(\xi)$ is finite in a neighborhood of zero and $M' := \mathbb{E}\eta(\xi) \leq 1+M$, where $M = \mathbb{E}\chi(\xi)$ is finite.*

*Proof.* Here we will show that conditions $(C1)$ and $(C2)$ hold true. We establish the validity of $(C3)$ later on, after presenting some lemmas. Take any $x \in X$ and $s \in \mathbb{R}$. Note that (A1) implies that

$$Y := F(x,\xi)-s \tag{3.3.61}$$

has finite expected value. Moreover,

$$Y-\mathbb{E}Y = (F(x,\xi)-s)-\mathbb{E}\left[F(x,\xi)-s\right] = F(x,\xi)-f(x) \tag{3.3.62}$$

is a $\sigma$-sub-Gaussian random variable by $(A2)$. Since $\phi$ is $L(\phi)$-Lipschitz continuous, we conclude by Proposition 2.4.5 that:

$$\phi(F(x,\xi)-s) \tag{3.3.63}$$

has finite expected value and that:

$$\phi(F(x,\xi)-s)-\mathbb{E}\phi(F(x,\xi)-s) \tag{3.3.64}$$

is a $(\kappa L(\phi)\sigma)$-sub-Gaussian random variable, where $\kappa \leq 6.86$ is an absolute constant. This shows that conditions (C1) and (C2) are satisfied. □

The following lemmas will be useful for showing that condition (C3) is also satisfied.

[13] See Proposition 2.4.5.

**Lemma 3.3.9.** *Define* $h : \mathbb{R}^{n+m} \to \mathbb{R}$ *as:*

$$h(z) = h(x, y) := \alpha||x||_2 + \beta||y||_2, \tag{3.3.65}$$

*where* $x \in \mathbb{R}^n$*,* $y \in \mathbb{R}^m$*;* $\alpha, \beta \in \mathbb{R}$ *and* $n, m \in \mathbb{N}$*. Then,* $|h(z)| \le \sqrt{\alpha^2 + \beta^2}||z||_2$*.*

*Proof.* We have that:

$$\begin{aligned} |h(z)| = |\alpha \left\|x\right\|_2 + \beta \left\|y\right\|_2| &= \left|\left\langle \begin{pmatrix} ||x||_2 \\ ||y||_2 \end{pmatrix}, \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \right\rangle\right| && (3.3.66) \\ &\le \sqrt{\alpha^2 + \beta^2}\sqrt{\left\|x\right\|_2^2 + \left\|y\right\|_2^2} && (3.3.67) \\ &= \sqrt{\alpha^2 + \beta^2}\left\|z\right\|_2, && (3.3.68) \end{aligned}$$

applying the Cauchy-Schwarz inequality. □

**Remark 3.3.10.** *We can suppose, without loss of generality, that the norm in equation (3.3.1) of assumption (A3) is the Euclidean one. In fact, let* $n \in \mathbb{N}$ *be given and take any norm* $\|\cdot\|$ *on* $\mathbb{R}^n$*. Let us suppose that assumption (A3) holds with this particular norm. We will show that these assumption is also satisfied with the Euclidean norm up to a suitable multiplicative constant. Indeed, since* $\mathbb{R}^n$ *is finite-dimensional, there exists a real constant* $D > 0$ *such that:*

$$\|x\| \le D \left\|x\right\|_2, \ \forall x \in \mathbb{R}^n. \tag{3.3.69}$$

*So, w.p.1* $\xi \in \operatorname{supp}(\xi)$*, we have that:*

$$|F(x', \xi) - F(x, \xi)| \le D\chi(\xi) \left\|x' - x\right\|_2, \tag{3.3.70}$$

*for all* $x', x \in X$*. Of course,* $M_{D\chi}(s) = M_\chi(Ds)$*, for all* $s \in \mathbb{R}$*. So, the moment generating function of* $D\chi(\xi)$ *is also finite in a neighborhood of zero.* □

From now on, except when otherwise stated, we will always use the Euclidean norm on $\mathbb{R}^n$, where $n \in \mathbb{N}$. We will also denote it just by $\|\cdot\|$, instead of $\|\cdot\|_2$.

**Lemma 3.3.11.** *Suppose that assumption (A3) holds true. Let* $\nu : \mathbb{R} \to \mathbb{R}$ *be a* $L-$*Lipschitz continuous function and consider the family of random variables:*

$$G(x, s, \xi) := \nu(F(x, \xi) - s), \tag{3.3.71}$$

*for all* $x \in X$ *and* $s \in \mathbb{R}$*. Then, w.p.1,*

$$|G(x', s', \xi) - G(x, s, \xi)| \le L\eta(\xi) \left\|(x, s) - (x', s')\right\|, \tag{3.3.72}$$

*for all* $x, x' \in X$ *and* $s, s' \in \mathbb{R}$*, where* $\eta(\xi)$ *is defined as in equation (3.3.59). Moreover, the random variable* $\eta(\xi)$ *satisfies:*

$$M = \mathbb{E}\chi(\xi) \le \mathbb{E}\eta(\xi) \le 1 + \mathbb{E}\chi(\xi) = 1 + M, \tag{3.3.73}$$

*and also* $\operatorname{dom} M_\eta = \operatorname{dom} M_\chi$*.*

*Proof.* By assumption (A3), there exists a measurable function $\chi : \mathrm{supp}(\xi) \to \mathbb{R}_+$ whose moment generating function is finite in a neighborhood of zero and that satisfies

$$|F(x',\xi) - F(x,\xi)| \leq \chi(\xi) \|x' - x\|, \tag{3.3.74}$$

for all $x', x \in X$ and $\xi \in E \subseteq \mathrm{supp}\,\xi$, where $\mathbb{P}[\xi \in E] = 1$. Take any $x', x \in X$ and $s', s \in \mathbb{R}$. So, for every $\xi \in E$, we have that:

$$\begin{aligned} |G(x',s',\xi) - G(x,s,\xi)| &\leq L\,|(F(x,\xi) - s) - (F(x',\xi) - s')| && (3.3.75)\\ &\leq L(|F(x,\xi) - F(x',\xi)| + |s - s'|) && (3.3.76)\\ &\leq L(\chi(\xi)\,\|x - x'\| + |s - s'|) && (3.3.77)\\ &\leq L\sqrt{\chi(\xi)^2 + 1}\,\|(x,s) - (x',s')\|, && (3.3.78) \end{aligned}$$

using Lemma 3.3.9 in the last inequality.

Since

$$z \leq \sqrt{z^2 + 1} \leq 1 + z, \ \forall z \geq 0, \tag{3.3.79}$$

and $\chi(\xi) \geq 0$, for all $\xi \in \mathrm{supp}(\xi)$, it follows that:

$$\chi(\xi) \leq \eta(\xi) \leq 1 + \chi(\xi), \ \forall \xi \in \mathrm{supp}(\xi). \tag{3.3.80}$$

So,

$$M = \mathbb{E}\chi(\xi) \leq \mathbb{E}\eta(\xi) \leq \mathbb{E}\left[1 + \chi(\xi)\right] = 1 + M. \tag{3.3.81}$$

Now, let us show that $\mathrm{dom}\, M_\eta = \mathrm{dom}\, M_\chi$. Since $\chi(\xi)$ and $\eta(\xi)$ are non-negative random variables, it follows that their moment generating functions are finite[14] for $s \leq 0$. Since $\eta(\xi) \geq \chi(\xi)$, we obtain that $M_\eta(s) \geq M_\chi(s)$, for all $s > 0$. This implies that $\mathrm{dom}\, M_\eta \subseteq \mathrm{dom}\, M_\chi$. Finally, note that:

$$M_\eta(s) = \mathbb{E}\exp\{s\eta(\xi)\} \leq \mathbb{E}\exp\{s(1 + \chi(\xi))\} = \exp(s)\, M_\chi(s) < +\infty, \tag{3.3.82}$$

for all $s > 0$ such that $s \in \mathrm{dom}\, M_\chi$. We conclude that $\mathrm{dom}\, M_\eta \supseteq \mathrm{dom}\, M_\chi$, which concludes the proof. ☐

Take any $\phi \in \Phi$ satisfying $L(\phi) < +\infty$. As a direct consequence of Lemma 3.3.11, assumption (A3) implies that the family of random variables

$$\{\phi\left(F(x,\xi) - s\right) : x \in X, s \in \mathbb{R}\}$$

satisfies condition (C3) (see Proposition 3.3.8). Indeed, we have shown that $L(\phi)\eta(\xi)$ has a finite moment generating function in a neighborhood of zero[15] and satisfies, with probability 1, $\xi \in \mathrm{supp}(\xi)$,

$$|\phi\left(F(x',\xi) - s'\right) - \phi\left(F(x,\xi) - s\right)| \leq L(\phi)\eta(\xi)\,\|(x',s') - (x,s)\|, \tag{3.3.83}$$

[14] In fact, less or equal than $\exp(0) = 1$.

[15] Take any constant $c \in \mathbb{R}$ and any random variable $Z$. We have that $M_{cZ}(s) = M_Z(cs)$, for all $s \in \mathbb{R}$. Since $0 \in \mathrm{int}\,\mathrm{dom}\, M_\chi = \mathrm{int}\,\mathrm{dom}\, M_\eta$, it follows that $0 \in \mathrm{int}\, M_{L(\phi)\eta}$.

for all $(x', s'), (x, s) \in X \times \mathbb{R}$. Moreover, writing $M' := \mathbb{E}\eta(\xi)$, we also have that:

$$\mathbb{E}L(\phi)\eta(\xi) = L(\phi)M' \leq L(\phi)(1 + M). \tag{3.3.84}$$

Finally, in order to close this discussion, let us observe that:

$$I_{L(\phi)\eta(\xi)}(2L(\phi)M') = I_{\eta(\xi)}(2M'), \tag{3.3.85}$$

since $L(\phi) \neq 0$. Similarly to the constant $\mathsf{m} \in (0, +\infty]$ appearing in equation (2.1.28), we will consider an appropriate constant:

$$\mathsf{m}' := I_{\eta(\xi)}(2M') \in (0, +\infty] \tag{3.3.86}$$

in our sample complexity estimate for stochastic programming problems with OCE risk measures.

In Proposition 3.3.12 we present a series of assertions that follow from assumptions (A1)-(A5), when $\phi \in \Phi$ satisfies $l(\phi) < 1 < L(\phi) < +\infty$. Before proceeding, let us introduce some notation. For $N \in \mathbb{N}$ and $\epsilon > 0$, define the event[16]:

$$E_N^{\epsilon}(\tilde{X}) := \left[\sup_{(x,s)\in\tilde{X}} |v(x, s) - \hat{v}_N(x, s)| < \epsilon\right]. \tag{3.3.87}$$

Note that $E_N^{\delta}(\tilde{X}) \subseteq E_N^{\epsilon}(\tilde{X})$, whenever $0 \leq \delta \leq \epsilon$.

**Proposition 3.3.12.** *Take any $\phi \in \Phi$ satisfying $l(\phi) < 1 < L(\phi) < +\infty$ and suppose that assumptions $(A1) - (A5)$ hold true. Define $\mathfrak{l}(\cdot)$, $\mathfrak{u}(\cdot)$, $a(\cdot)$, $b(\cdot)$ and $\tilde{X}$, respectively, as in equations (3.3.54), (3.3.55), (3.3.21), (3.3.22) and (3.3.6). Then, the following assertions hold:*

(a) *$\tilde{X}$ is a nonempty compact set that satisfies:*

  (a1) *For every $x \in X$, $\operatorname{int} \tilde{X}_x = (a(x), b(x)) \supseteq \operatorname{argmin}_{s\in\mathbb{R}} v(x, s) \neq \emptyset$. In particular, $v^* = v^*(\tilde{X})$ holds true.*

  (a2) *Take any $N \in \mathbb{N}$. For almost every sample realization $\{\xi^1, \ldots, \xi^N\}$ of $\xi$, $\operatorname{argmin}_{s\in\mathbb{R}} \hat{v}_N(x, s) \neq \emptyset$, for all $x \in X$.*

  (a3) *$D' := \operatorname{diam} \tilde{X} \leq \sqrt{D^2 + (b - a)^2}$, where:*

$$\begin{aligned} a := \inf_{x\in X} a(x) &= \inf_{x\in X} f(x) + \mathfrak{l}\left(\frac{\gamma_1}{2}\right) - \overline{z} > -\infty, \textit{ and} \quad (3.3.88)\\ b := \sup_{x\in X} b(x) &= \sup_{x\in X} f(x) + \mathfrak{u}\left(\frac{1 + \gamma_2}{2}\right) - \underline{z} < +\infty. \quad (3.3.89)\end{aligned}$$

[16] Of course, these event also depends on $\phi \in \Phi$ and $\tilde{X} \subseteq \mathbb{R}^{n+1}$.

(*b*) *The objective functions of problems (3.2.1) and (3.2.7) are Lipschitz continuous on* $X$ *and on* $X \times \mathbb{R}$*, respectively, with Lipschitz constants equal to* $L(\phi)M$ *and* $(1 + L(\phi)M')$*. In particular, problems (3.2.1), (3.2.7) and (3.2.27) are solvable, i.e.* $S \neq \emptyset$, $ES \neq \emptyset$ *and* $ES(\tilde{X}) \neq \emptyset$*. Moreover,* $ES = ES(\tilde{X})$ *holds true.*

(*c*) *Take any* $N \in \mathbb{N}$*. For almost every sample realization* $\{\xi^1, \ldots, \xi^N\}$ *of* $\xi$*, we have that* $\hat{v}_N : X \to \mathbb{R}$ *and* $\hat{v}_N : X \times \mathbb{R} \to \mathbb{R}$ *are Lipschitz continuous functions on* $X$ *and on* $X \times \mathbb{R}$*, respectively, with Lipschitz constants equal to:*

$$\frac{L(\phi)}{N} \sum_{i=1}^{N} \chi(\xi^i) \text{ and } \left(1 + \frac{L(\phi)}{N} \sum_{i=1}^{N} \eta(\xi^i)\right). \tag{3.3.90}$$

*In particular, we conclude that:*

$$\mathbb{P}\left[\hat{S}_N \neq \emptyset\right] = \mathbb{P}\left[\hat{ES}_N \neq \emptyset\right] = \mathbb{P}\left[\hat{ES}_N(\tilde{X}) \neq \emptyset\right] = 1. \tag{3.3.91}$$

(*d*) *Define* $\Delta(x)$ *as in equation (3.3.7), for all* $x \in X$*. It follows that:*

$$\Delta := \inf_{x \in X} \Delta(x) > 0, \tag{3.3.92}$$

*Moreover, whenever the event* $E_N^{\Delta}(\tilde{X})$ *happens,*

$$\operatorname{int} \tilde{X}_x \supseteq \{s \in \mathbb{R} : \hat{v}_N(x,s) \leq \hat{v}_N^* + \Delta\} \supseteq \operatorname*{argmin}_{s \in \mathbb{R}} \hat{v}_N(x,s),\ \forall x \in X. \tag{3.3.93}$$

*In that case, it is also true that:*

$$\hat{ES}_N = \hat{ES}_N(\tilde{X}). \tag{3.3.94}$$

*Proof.* Let us begin by proving item (a). Let us recall that

$$\tilde{X} = \{(x,s) \in X \times \mathbb{R} : a(x) \leq s \leq b(x)\} \subseteq \mathbb{R}^{n+1}. \tag{3.3.95}$$

Since $\phi \in \Phi$ satisfies $l(\phi) < 1 < L(\phi) < +\infty$ and assumptions (A1) and (A2) hold, we conclude from Proposition 3.1.17 and Corollary 3.3.6 that:

$$\begin{aligned}
\emptyset \neq \operatorname*{argmin}_{s \in \mathbb{R}} v(x,s) &\subseteq [f(x) + \mathfrak{l}(\gamma_1) - \overline{z}, f(x) + \mathfrak{u}(\gamma_2) - \underline{z}] \\
&\subseteq (f(x) + \mathfrak{l}(\gamma_1/2) - \overline{z}, f(x) + \mathfrak{u}((1+\gamma_2)/2) - \underline{z}) \\
&= (a(x), b(x)) = \operatorname{int} \tilde{X}_x.
\end{aligned}$$

Now, it follows from Proposition 3.2.1 that $v^* = v^*(\tilde{X})$, which proves item (a1). Since assumption (A3) is also satisfied, Proposition 2.1.4 implies that $f : X \to \mathbb{R}$ is Lipschitz continuous. We conclude that both functions $a, b : X \to \mathbb{R}$ are Lipschitz continuous on $X$. Let us recall that $X$ is a nonempty compact set (assumption (A4)). Then,

$$-\infty < a = \inf_{x\in X} f(x) + \mathfrak{l}\left(\frac{\gamma_1}{2}\right) - \overline{z} < \sup_{x\in X} f(x) + \mathfrak{u}\left(\frac{1+\gamma_2}{2}\right) - \underline{z} = b < +\infty. \quad (3.3.96)$$

It follows that:

$$\tilde{X} \subseteq X \times [a,b] \quad (3.3.97)$$

is a nonempty compact set (by the compactness and non-emptiness of $X$ and the continuity[17] of $a, b : X \to \mathbb{R}$) whose diameter is bounded by:

$$D' = \operatorname{diam} \tilde{X} \leq \operatorname{diam}(X \times [a,b]) = \sqrt{D^2 + (b-a)^2}. \quad (3.3.98)$$

This shows item $(a3)$. Now let us define the following event:

$$\mathcal{E} := \bigcap_{i\in\mathbb{N}} \left[\xi^i \in E\right] \in \mathcal{F}, \quad (3.3.99)$$

where $E$ is a Borel-measurable set of $\mathbb{R}^d$ satisfying $\mathbb{P}[\xi \in E] = 1$ and equation (2.1.11) (see Remark 2.1.1). Since $\xi^i \overset{d}{\sim} \xi$, for all $i \in \mathbb{N}$, it follows that:

$$\mathbb{P}(\mathcal{E}) = 1. \quad (3.3.100)$$

When the event $\mathcal{E}$ happens, we have that each function:

$$x \in X \mapsto F(x, \xi^i),\ i = 1, \ldots, N, \quad (3.3.101)$$

is $\chi(\xi^i)$-Lipschitz continuous on X. In particular, $F(\cdot, \xi^i)$ is bounded on $X$, for all $i = 1, \ldots, N$. So, there exist real numbers (depending on the sample realization) $\hat{u}_N \leq \hat{U}_N$ such that:

$$\hat{u}_N \leq F(x, \xi^i) \leq \hat{U}_N, \quad (3.3.102)$$

for all $x \in X$ and $i = 1, \ldots, N$. In particular, $F(x, \xi^i)$ is finite, for all $x \in X$ and $i = 1, \ldots, N$. So, Lemma 3.3.1 implies that $\hat{v}_N(x, \cdot)$ is a finite-valued convex function that satisfies:

$$\begin{aligned} \emptyset \neq \operatorname*{argmin}_{s\in\mathbb{R}} \hat{v}_N(x,s) &\subseteq \left[q^-_{\gamma_1}(F(x,\hat{\xi})) - \overline{z}, q^+_{\gamma_2}(F(x,\hat{\xi})) - \underline{z}\right] && (3.3.103)\\ &\subseteq \left[\hat{u}_N - \overline{z}, \hat{U}_N - \underline{z}\right]. && (3.3.104) \end{aligned}$$

This finishes the proof of item (a)[18].

[17] Note also that $a(x) < b(x)$, for all $x \in X$.

[18] The assertion "$\forall x \in X : \operatorname{argmin}_{s\in\mathbb{R}} \hat{v}_N(x,s) \neq \emptyset$" was the only fact needed for finishing the proof of item (a2). Nevertheless, we use later that these solution sets are all contained in the bounded interval $\left[\hat{u}_N - \overline{z}, \hat{U}_N - \underline{z}\right]$.

Now, let us show item (b). Take any $x, x' \in X$. Observe that:

$$\begin{aligned}
|v(x') - v(x)| &= \left|\inf_{s\in\mathbb{R}} v(x', s) - \inf_{s\in\mathbb{R}} v(x, s)\right| && (3.3.105)\\
&\leq \sup_{s\in\mathbb{R}} |v(x', s) - v(x, s)| && (3.3.106)\\
&= \sup_{s\in\mathbb{R}} |\mathbb{E}\phi(F(x', \xi) - s) - \mathbb{E}\phi(F(x, \xi) - s)| && (3.3.107)\\
&\leq \sup_{s\in\mathbb{R}} L(\phi)\mathbb{E}\, |F(x', \xi) - F(x, \xi)| && (3.3.108)\\
&\leq L(\phi)\mathbb{E}\chi(\xi) \|x' - x\| && (3.3.109)\\
&= L(\phi)M \|x' - x\| . && (3.3.110)
\end{aligned}$$

Note that equations (3.3.106) and (3.3.109) follow, respectively, from Proposition 2.8.4 and assumption (A3). Since $X$ is nonempty compact, we conclude that $S \neq \emptyset$. Moreover, it follows from Proposition 3.2.1 that $S = \pi_x\left(ES(\tilde{X})\right)$, so $ES(\tilde{X}) \neq \emptyset$[19]. We also have that $ES = ES(\tilde{X})$. Indeed, take any $(\bar{x}, \bar{s}) \in ES$. Then,

$$v(\bar{x}, \bar{s}) = v^* \leq v(\bar{x}, s),\ \forall s \in \mathbb{R}, \tag{3.3.111}$$

i.e. $\bar{s} \in \operatorname{argmin}_{s\in\mathbb{R}} v(\bar{x}, s) \subseteq \tilde{X}_{\bar{x}}$. So, $(\bar{x}, \bar{s}) \in \tilde{X}$. The converse inclusion is trivially true, since $v^*(\tilde{X}) = v^*$.

Now, we show that $v : X \times \mathbb{R} \to \mathbb{R}$ is Lipschtiz continuous. By Proposition 3.3.8 it follows that:

$$|\phi(F(x', \xi) - s') - \phi(F(x, \xi) - s)| \leq L(\phi)\eta(\xi) \|(x', s') - (x, s)\| , \tag{3.3.112}$$

for all $(x', s'), (x, s) \in X \times \mathbb{R}$ and $\xi \in E$[20]. Therefore,

$$\begin{aligned}
|v(x', s') - v(x, s)| &\leq |s' - s| + |\mathbb{E}\phi(F(x', \xi) - s') - \mathbb{E}\phi(F(x, \xi) - s)| && (3.3.113)\\
&\leq |s' - s| + \mathbb{E}\, |\phi(F(x', \xi) - s') - \phi(F(x, \xi) - s)| && (3.3.114)\\
&\leq |s' - s| + L(\phi)\mathbb{E}\eta(\xi) \|(x', s') - (x, s)\| && (3.3.115)\\
&= |s' - s| + L(\phi)M' \|(x', s') - (x, s)\| && (3.3.116)\\
&\leq (1 + L(\phi)M') \|(x', s') - (x, s)\| , && (3.3.117)
\end{aligned}$$

which finishes the proof of item (b).

Now, let us show item (c). Following the same reasoning used in the proof of item (b), it is elementary to show that, whenever the event $\mathcal{E}$ happens,

$$|\hat{v}_N(x') - \hat{v}_N(x)| \leq L(\phi)\hat{\mathbb{E}}\chi(\hat{\xi}) \|x' - x\| = \frac{L(\phi)}{N}\sum_{i=1}^{N} \chi(\xi^i) \|x' - x\| , \tag{3.3.118}$$

[19] This is a somewhat indirect way for proving this fact. An alternative way would be to argue that $v : X \times \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous and $\tilde{X}$ is nonempty compact.

[20] This is the same set considered before in the proof (see also Remark 2.1.1).

for all $x', x \in X$, and

$$\begin{aligned}|\hat{v}_N(x', s') - \hat{v}_N(x, s)| &\leq \left(1 + L(\phi)\hat{\mathbb{E}}\eta(\hat{\xi})\right) \|(x', s') - (x, s)\| & (3.3.119)\\ &= \left(1 + \frac{L(\phi)}{N}\sum_{i=1}^{N}\eta(\xi^i)\right) \|(x', s') - (x, s)\|, & (3.3.120)\end{aligned}$$

for all $x', x \in X$ and $s', s \in \mathbb{R}$. The compactness of $X$ and $\tilde{X}$ imply that $\hat{S}_N \neq \emptyset$ and $\hat{ES}_N(\tilde{X}) \neq \emptyset$, whenever $\mathcal{E}$ happens, i.e.

$$\mathbb{P}\left[\hat{S}_N \neq \emptyset\right] = \mathbb{P}\left[\hat{ES}_N \neq \emptyset\right] = 1. \qquad (3.3.121)$$

Whenever the event $\mathcal{E}$ happens, the inclusion (3.3.104) guarantees that $\hat{ES}_N = \hat{ES}_N\left(X \times \left[\hat{u}_N, \hat{U}_N\right]\right)$. Finally, observe that $\hat{ES}_N\left(X \times \left[\hat{u}_N, \hat{U}_N\right]\right) \neq \emptyset$, since $X \times \left[\hat{u}_N - \overline{z}, \hat{U}_N - \underline{z}\right]$ is (nonempty) compact. This finishes the proof of item (c).

Now, we prove item (d). We first note that $\Delta$ is positive by Corollary 3.3.7, since assumptions (A1) and (A2) are satisfied. Let us assume, without loss of generality, that the event $E_N^{\Delta}(\tilde{X}) \cap \mathcal{E}$ happens[21]. Take any $x \in X$. Since the event $E_N^{\Delta}(\tilde{X})$ happens, it follows that:

$$\sup_{s \in [a(x), b(x)]} |\hat{v}_N(x, s) - v(x, s)| < \Delta \leq \Delta(x). \qquad (3.3.122)$$

From item (a), there exists $s(x) \in \operatorname{argmin}_{s\in\mathbb{R}} v(x, s) \subseteq (a(x), b(x))$, so

$$|\hat{v}_N(x, s(x)) - v(x)| < \Delta(x).$$

It follows that equation (3.3.8) is satisfied. We also have that $F(x, \xi^i)$ is finite, for all $i = 1, \ldots, N$, since the event $\mathcal{E}$ happens. Proposition 3.3.2 implies that:

$$\begin{aligned}\operatorname{int}\tilde{X}_x &\supseteq \{\hat{v}_N(x, s) \leq \hat{v}_N(x) + \Delta(x)\} & (3.3.123)\\ &\supseteq \{\hat{v}_N(x, s) \leq \hat{v}_N(x) + \Delta\} & (3.3.124)\\ &\supseteq \operatorname*{argmin}_{s\in\mathbb{R}} \hat{v}_N(x, s). & (3.3.125)\end{aligned}$$

Finally, let us show that $\hat{ES}_N = \hat{ES}_N(\tilde{X})$, whenever $E_N^{\Delta}(\tilde{X}) \cap \mathcal{E}$ happens. Take any $(\bar{x}, \bar{s}) \in \hat{ES}_N$. Note that $\bar{s} \in \operatorname{argmin}_{s\in\mathbb{R}} v(\bar{x}, s) \subseteq \tilde{X}_x$, i.e. $(\bar{x}, \bar{s}) \in \tilde{X}$. Since $\hat{v}_N \leq \hat{v}_N(\tilde{X})$, it follows that $(\bar{x}, \bar{s}) \in \hat{ES}_N(\tilde{X})$. For showing the converse inclusion, note that Proposition 3.2.1 implies that $\hat{v}_N(\tilde{X}) = \hat{v}_N$. The proof of the proposition is complete. □

[21] The events $E_N^{\Delta}(\tilde{X}) \cap \mathcal{E}$ and $E_N^{\Delta}(\tilde{X})$ have equal probabilities.

**Remark 3.3.13.** *Let us obtain a more explicit bound for $D' = \operatorname{diam} \tilde{X}$ in terms of $D$, $M$, $L(\phi)$, $l(\phi)$, $\overline{z}$ and $\underline{z}$. Since $X$ is compact and $f : X \to \mathbb{R}$ is continuous, there exist $\overline{x} \in X$ and $\underline{x} \in X$ such that*

$$\sup_{x\in X} f(x) - \inf_{x\in X} f(x) = f(\overline{x}) - f(\underline{x}) \leq M \|\overline{x} - \underline{x}\| \leq MD. \tag{3.3.126}$$

*Therefore,*

$$\begin{aligned}
D' &\leq \sqrt{D^2 + (b-a)^2} \\
&\leq D + b - a \\
&= D + \sup_{x\in X} f(x) - \inf_{x\in X} f(x) + \sigma \left[\sqrt{2\log\left(\frac{2}{1-\gamma_2}\right)} + \sqrt{2\log\left(\frac{2}{\gamma_1}\right)}\right] + \overline{z} - \underline{z} \\
&\leq (M+1)\,D + \overline{z} - \underline{z} + \sigma \left[\sqrt{2\log\left(\frac{2(L(\phi) - 1 + L(\phi) - l(\phi))}{1 - l(\phi)}\right)} + \right. \\
&\qquad \left. \sqrt{2\log\left(\frac{2(L(\phi) - l(\phi) + 1 - l(\phi))}{L(\phi) - 1}\right)}\right].
\end{aligned}$$

□

In the remainder of this section we will assume that $\mathfrak{l}(\cdot)$, $\mathfrak{u}(\cdot)$, $a(\cdot)$, $b(\cdot)$ and $\tilde{X}$ are defined, respectively, as in equations (3.3.54), (3.3.55), (3.3.21), (3.3.22) and (3.3.6). Moreover, the event $\mathcal{E}$ is defined as in equation (3.3.99) and:

$$\begin{aligned}
\Delta &:= \inf_{x\in X} \Delta(x), \\
a &:= \inf_{x\in X} a(x), \\
b &:= \sup_{x\in X} b(x).
\end{aligned}$$

We are ready to establish the main result of this section.

**Proposition 3.3.14.** *Take any $\phi \in \Phi$ satisfying $l(\phi) < 1 < L(\phi) < +\infty$. Suppose that assumptions $(A1) - (A5)$ hold true. Then, for any $\epsilon > 0$,*

$$\mathbb{P}\left(E_N^{\epsilon}(\tilde{X})\right) \;\geq 1 - \exp\{-N\mathsf{m}'\} - 2\left[\frac{4\rho D' M'}{\epsilon}\right]^{n+1} \exp\left\{-\frac{N\epsilon^2}{32C^2 L(\phi)^2\sigma^2}\right\}, \tag{3.3.127}$$

*where $\eta = \sqrt{1+\chi(\xi)^2}$; $M' = \mathbb{E}\eta(\xi)$, $\mathsf{m}' = I_\eta(2M') \in (0, +\infty]$, $D' = \operatorname{diam} \tilde{X}$ and $\sigma^2$ are constants depending on the problem data; and $\kappa$ and $\rho$ are absolute constants. Moreover, if $0 < \epsilon \leq \Delta$, then*

$$\begin{aligned}
\mathbb{P}\left[\sup_{x\in X} |\hat{v}_N(x) - v(x)| < \epsilon\right] &\geq 1 - \exp\{-N\mathsf{m}'\} \\
&\quad -2\left[\frac{4\rho D' M'}{\epsilon}\right]^{n+1} \exp\left\{-\frac{N\epsilon^2}{32C^2 L(\phi)^2\sigma^2}\right\},
\end{aligned} \tag{3.3.128}$$

*Proof.* Take any $\epsilon > 0$. Since assumptions (A1)-(A3) hold true, Proposition 3.3.8 implies that conditions (C1)-(C3) are satisfied by the family of random variables $\{\phi(F(x,\xi)-s) : x \in X, s \in \mathbb{R})\}$. Moreover, $\tilde{X}$ is a nonempty compact set (item (a) of Proposition 3.3.12) and assumption (A5) holds true. So, we can apply Theorem 2.1.5, considering the family of random variables $\{\phi(F(x,\xi)-s) : (x,s) \in \tilde{X}\}$, in order to obtain the lower bound in equation (3.3.127), recalling that:

$$E_N^\epsilon(\tilde{X}) = \left[\sup_{(x,s)\in\tilde{X}} |\hat{v}_N(x,s) - v(x,s)| < \epsilon\right]. \tag{3.3.129}$$

In fact, equation (3.3.127) is similar to equation (2.1.28)[22] with the parameters $n+1$, $D'$, $M'$, $m'$ and $\kappa^2 L(\phi)^2\sigma^2$ playing the role, respectively, of $n$, $D$, $M$, $m$ and $\sigma^2$.

Now, take $0 < \epsilon \leq \Delta$. Item (d) of Proposition 3.3.12 guarantees that $\Delta > 0$. We also have that

$$\mathbb{P}\left(\mathcal{E} \cap E_N^\epsilon(\tilde{X})\right) = \mathbb{P}\left(E_N^\epsilon(\tilde{X})\right),$$

since the event $\mathcal{E}$ has probability 1. For finishing the proof, we just need to show that

$$\mathcal{E} \cap E_N^\epsilon(\tilde{X}) \subseteq \left[\sup_{x\in X} |\hat{v}_N(x) - v(x)| < \epsilon\right]. \tag{3.3.130}$$

Since $\epsilon \leq \Delta$, $E_N^\epsilon(\tilde{X}) \subseteq E_N^\Delta(\tilde{X})$. Item (d) of Proposition 3.3.12 implies that $\tilde{X}_x \supseteq \operatorname{argmin}_{s\in\mathbb{R}} \hat{v}_N(x,s) \neq \emptyset$, for all $x \in X$, whenever the event $\mathcal{E} \cap E_N^\epsilon(\tilde{X})$ happens. We also have by item (a1) of 3.3.12 that $\tilde{X}_x \supseteq \operatorname{argmin}_{s\in\mathbb{R}} v(x,s) \neq \emptyset$, for all $x \in X$. So, whenever the event $E_N^\epsilon(\tilde{X}) \cap \mathcal{E}$ happens, we have that

$$\begin{aligned} |\hat{v}_N(x) - v(x)| &= \left|\inf_{s\in\tilde{X}_x} \hat{v}_N(x,s) - \inf_{s\in\tilde{X}_x} v(x,s)\right| && (3.3.131)\\ &\leq \sup_{s\in\tilde{X}_x} |\hat{v}_N(x,s) - v(x,s)| && (3.3.132)\\ &\leq \sup_{(x,s)\in\tilde{X}} |\hat{v}_N(x,s) - v(x,s)| < \epsilon, && (3.3.133)\end{aligned}$$

Therefore,

$$\mathcal{E} \cap E_N^\epsilon(\tilde{X}) \subseteq \left[\sup_{x\in X} |\hat{v}_N(x) - v(x)| < \epsilon\right],$$

and equation (3.3.128) holds true. □

**Remark 3.3.15.** *It worth making two points regarding the validity of the exponential bound (3.3.128) for $\epsilon > 0$ sufficiently small. Equation (3.3.128) was shown for*

[22] From Theorem 2.1.5 we obtain a upper bound for $\mathbb{P}\left(E_N^\epsilon(\tilde{X})^C\right) = 1 - \mathbb{P}\left(E_N^\epsilon(\tilde{X})\right)$. Of course, the lower bound in equation (3.3.127) follows immediately from that upper bound.

$0 < \epsilon \leq \Delta$, *where*

$$\begin{array}{rl} \Delta & := \min_{x\in X} \Delta(x) \\ & \geq \frac{1}{3}\min\left\{\frac{L(\phi)-1}{8}\left[\mathfrak{l}(\frac{3\gamma_1}{4}) - \mathfrak{l}(\frac{\gamma_1}{2}))\right], \frac{1-l(\phi)}{8}\left[\mathfrak{u}(\frac{1+\gamma_2}{2}) - \mathfrak{u}(\frac{1}{4}+\frac{3\gamma_2}{4})\right]\right\} > 0, \end{array} \tag{3.3.134}$$

*by Proposition 3.3.4. Of course, one can obtain an exponential bound like (3.3.128) for* $\epsilon > \Delta$, *since, in that case, we have that*

$$\begin{array}{rl} \mathbb{P}\left[\sup_{x\in X}|\hat{v}_N(x) - v(x)| < \epsilon\right] & \geq \mathbb{P}\left[\sup_{x\in X}|\hat{v}_N(x) - v(x)| < \Delta\right] \\ & \geq 1 - \exp\{-N\mathfrak{m}'\} - C\exp\{-N\beta(\Delta)\}, \end{array} \tag{3.3.135}$$

*for all* $N \in \mathbb{N}$, *where*

$$\beta(\Delta) := \frac{\Delta^2}{32\kappa^2 L(\phi)^2\sigma^2} > 0. \tag{3.3.136}$$

*Second, since* $F(x,\xi) - f(x)$ *is* $\sigma$*-sub-Gaussian, for every* $x \in X$, *we can take*

$$\mathfrak{l}(\gamma) := -\sigma\sqrt{2\log\left(\frac{1}{\gamma}\right)}, \forall \gamma \in (0,1), \text{ and} \tag{3.3.137}$$

$$\mathfrak{u}(\gamma) := \sigma\sqrt{2\log\left(\frac{1}{1-\gamma}\right)}, \forall \gamma \in (0,1), \tag{3.3.138}$$

*in (3.3.134). Note also that* $\mathfrak{l}(\gamma) \to -\infty$, *when* $\gamma \to 0$, *and* $\mathfrak{u}(\gamma) \to \infty$, *when* $\gamma \to 1$. *In particular, by taking* $0 < \gamma_1' < \gamma_1/2$ *sufficiently small and* $\frac{1+\gamma_2}{2} < \gamma_2' < 1$ *sufficiently large and by defining* $a(x)$ *and* $b(x)$ *as*

$$a(x) := f(x) + \mathfrak{l}(\gamma_1') - \overline{z}, \text{ and} \tag{3.3.139}$$

$$b(x) := f(x) + \mathfrak{u}(\gamma_2') - \underline{z}, \tag{3.3.140}$$

*we can make* $\Delta$ *arbitrarily large. In that case, we still have that* $\tilde{X} = \{(x,s) \in X \times \mathbb{R} : a(x) \leq s \leq b(x)\}$ *is compact, although its diameter* $D'$ *gets larger with these choices of* $\gamma_1' < \gamma_1/2$ *and* $\gamma_2' > (1+\gamma_2)/2$. *This impacts the value of the constant*

$$C := 2\left[\frac{4\rho D' M'}{\epsilon}\right]^{n+1} \tag{3.3.141}$$

*in the exponential bound (3.3.128). Either way, we see that it is possible to obtain an exponential bound like (3.3.128) for arbitrarily large values of* $\epsilon > 0$. □

**Corollary 3.3.16.** *Take* $\phi \in \Phi$ *satisfying* $l(\phi) < 1 < L(\phi) < +\infty$. *Consider stochastic programming problems such as (3.2.1) and (3.2.7) and suppose that assumptions (A1)-(A5) hold true. Then, for* $\epsilon > 0$ *sufficiently small and for all* $0 \leq \delta < \epsilon$,

$$\begin{array}{rl} \mathbb{P}\left(\left[\hat{S}_N^\delta \subseteq S^\epsilon\right] \cap \left[\hat{S}_N^\delta \neq \emptyset\right]\right) \geq & 1 - \exp\{-N\mathsf{m}'\} \\ & -2\left[\frac{8\rho D' M'}{\epsilon-\delta}\right]^{n+1}\exp\left\{-\frac{N(\epsilon-\delta)^2}{128\kappa^2 L(\phi)^2\sigma^2}\right\}, \end{array} \tag{3.3.142}$$

*and,*

$$\mathbb{P}\left(\left[\hat{ES}_N^{\delta} \subseteq ES^{\epsilon}\right] \cap \left[\hat{ES}_N^{\delta} \neq \emptyset\right]\right) \geq \begin{array}{l} 1 - \exp\left\{-N\mathsf{m}'\right\} \\ -2\left[\frac{8\rho D'M'}{\epsilon-\delta}\right]^{n+1} \exp\left\{-\frac{N(\epsilon-\delta)^2}{128C^2L(\phi)^2\sigma^2}\right\}, \end{array} \tag{3.3.143}$$

*where $D'$, $M'$, $\mathsf{m}' \in (0, +\infty]$ and $\sigma^2$ are constants depending on the problem data; and $\kappa$ and $\rho$ are absolute constants.*

*Proof.* Take any $\epsilon \leq 2\Delta$ and $0 \leq \delta < \epsilon$. It follows that:

$$0 < \frac{\epsilon - \delta}{2} \leq \Delta.$$

From Proposition 3.3.14, it follows that:

$$\mathbb{P}\left(E_N^{\frac{\epsilon-\delta}{2}}(\tilde{X})\right) \geq 1 - \exp\left\{-N\mathsf{m}'\right\} - 2\left[\frac{8\rho D'M'}{\epsilon}\right]^{n+1} \exp\left\{-\frac{N\left(\epsilon-\delta\right)^2}{128\kappa^2L(\phi)^2\sigma^2}\right\} \tag{3.3.144}$$

is satisfied. Moreover, the events $E_N^{\frac{\epsilon-\delta}{2}}(\tilde{X})$ and $\mathcal{E} \cap E_N^{\frac{\epsilon-\delta}{2}}(\tilde{X})$ have the same probability. For finishing the proof we just need to show that $\mathcal{E} \cap E_N^{\frac{\epsilon-\delta}{2}}(\tilde{X})$ is contained in the events:

$$\left[\hat{ES}_N^{\delta} \subseteq ES^{\epsilon}\right] \cap \left[\hat{ES}_N^{\delta} \neq \emptyset\right] \text{ and } \left[\hat{S}_N^{\delta} \subseteq S^{\epsilon}\right] \cap \left[\hat{S}_N^{\delta} \neq \emptyset\right]. \tag{3.3.145}$$

In item (c) of Proposition 3.3.12 we have shown that $\mathcal{E}$ is contained on the events:

$$\left[\hat{ES}_N \neq \emptyset\right] \text{ and } \left[\hat{S}_N \neq \emptyset\right]. \tag{3.3.146}$$

In particular, $\mathcal{E}$ is contained on the events

$$\left[\hat{ES}_N^{\delta} \neq \emptyset\right] \text{ and } \left[\hat{S}_N^{\delta} \neq \emptyset\right]. \tag{3.3.147}$$

Now, note that item (i) and (ii) of Proposition 3.2.1 are satisfied. So, whenever the event $\mathcal{E} \cap E_N^{\frac{\epsilon-\delta}{2}}(\tilde{X})$ happens, we have that $\hat{ES}_N^{\delta}(\tilde{X}) \subseteq ES^{\epsilon}(\tilde{X})$ and $\hat{S}_N^{\delta} \subseteq S^{\epsilon}$. This already shows that

$$\mathcal{E} \cap E_N^{\frac{\epsilon-\delta}{2}}(\tilde{X}) \subseteq \left(\left[\hat{S}_N^{\delta} \subseteq S^{\epsilon}\right] \cap \left[\hat{S}_N \neq \emptyset\right]\right). \tag{3.3.148}$$

Finally, note that $ES = ES(\tilde{X})$ by item (b) of Proposition 3.3.12. Moreover, item (d) of the same proposition establishes that $\hat{ES}_N^{\delta} = \hat{ES}_N^{\delta}(\tilde{X})$ whenever the event $\mathcal{E} \cap E_N^{\Delta}(\tilde{X})$ happens. The result follows since $(\epsilon - \delta)/2 \leq \Delta$. □

**Remark 3.3.17.** *As we have shown in Proposition 3.1.11, $\mu_\phi(Z) = \mathbb{E}Z$, for every $Z \in \mathcal{Z}$, whenever $\phi \in \Phi$ satisfies either $l(\phi) = 1$ or $L(\phi) = 1$. In that case, we are back to the risk neutral setting, where equation (3.3.128) holds with:*

$$m' = I_\chi(2M), \tag{3.3.149}$$
$$M' = M, \tag{3.3.150}$$
$$D' = D, \tag{3.3.151}$$
$$\kappa = L(\phi) = 1. \tag{3.3.152}$$

*So, instead of supposing that $l(\phi) < 1 < L(\phi) < \infty$ in Proposition 3.3.14 in order to derive the exponential rate of convergence in (3.3.128), we can just suppose that $L(\phi) < \infty$.* □

Note that the sample complexity estimate obtained for risk averse stochastic programming problems with OCE risk measures is similar to the one obtained for risk neutral problems under the same regularity conditions. The main difference is that the constants appearing in the risk averse setting $M'$, $D'$, $\kappa$ and $L(\phi)$ typically slows down the exponential rate at which the probability of the desirable event (1.0.14) approaches 1 with the increase of the sample size $N$. For example, just considering the second exponential term in the sample complexity estimates in both risk neutral and risk averse settings, we obtain, respectively, the following sample sizes estimates

$$N_{rn} \geq \frac{128\sigma^2}{\epsilon^2}\left[n \log\left(\frac{4\rho D M}{\epsilon}\right) + \log\left(\frac{2}{\theta}\right)\right] \tag{3.3.153}$$

and

$$N_{ra} \geq \frac{128\kappa^2 L(\phi)^2\sigma^2}{\epsilon^2}\left[(n+1)\log\left(\frac{4\rho D' M'}{\epsilon}\right) + \log\left(\frac{2}{\theta}\right)\right]. \tag{3.3.154}$$

So, if[23] $L(\phi) = 10$ and assuming that the absolute constant $\kappa$ is equal to 6, the sample complexity in the risk averse setting is at least $3,600$ greater than the sample complexity in the risk neutral setting. We see in the next chapter that the situation gets even worse in the risk averse multistage setting.

[23] For example, $\mu_\phi(\cdot)$ is the $\mathrm{AV@R}_{1-\alpha}(\cdot)$ risk measure with $\alpha = 0.9$. It is worth mentioning that in the field of risk management one usually considers greater values for $\alpha$, like 0.95 or even 0.99.

# CHAPTER 4

## Sample complexity for dynamic problems with OCE risk measures

In this chapter we derive sample complexity estimates for a class of risk averse multistage stochastic programming problems. As before, we approximate the true problem by constructing a scenario tree via a Monte Carlo sampling scheme. In order to formulate a risk averse multistage stochastic programming problem with OCE risk measures, we begin by defining conditional OCE risk measures.

**Definition 4.0.18.** *Let $\mathcal{Z} = L_p(\Omega, \mathcal{F}, \mathbb{P})$, for some $p \in [1, \infty)$. Let $\phi \in \Phi$ be such that $\phi(Z) \in L_p(\Omega, \mathcal{F}, \mathbb{P})$, for every $Z \in \mathcal{Z}$. Given a sub-$\sigma$-algebra $\mathcal{G} \subseteq \mathcal{F}$ we define the conditional OCE risk measure $\mu_{\phi|\mathcal{G}} : \mathcal{Z} \to L_p(\Omega, \mathcal{G}, \mathbb{P})$ as*

$$\mu_{\phi|\mathcal{G}}(Z)(\omega) := \inf_{Y \in L_p(\Omega, \mathcal{G}, \mathbb{P})} \left\{ Y(\omega) + \mathbb{E}\left[\phi\left(Z - Y\right)|\mathcal{G}\right](\omega) \right\}. \tag{4.0.1}$$

**Remark 4.0.19.** *Note that if $L(\phi) < \infty$, then $\phi(Z) \in L_p(\Omega, \mathcal{F}, \mathbb{P})$, for every $Z \in L_p(\Omega, \mathcal{F}, \mathbb{P})$. Indeed, this is an immediate consequence of the inequality*

$$|\phi(z)| \leq L(\phi)\,|z|\,,\ \forall z \in \mathbb{R}, \tag{4.0.2}$$

*since $\phi(0) = 0$.* □

The following proposition shows that an OCE conditional risk measure is a convex conditional risk measure.

**Proposition 4.0.20.** *Let $\mathcal{Z} = L_p(\Omega, \mathcal{F}, \mathbb{P})$, for some $p \in [1, \infty)$. Let $\phi \in \Phi$ be such that $\phi(Z) \in L_p(\Omega, \mathcal{F}, \mathbb{P})$, for every $Z \in \mathcal{Z}$. Take any sub-$\sigma$-algebra $\mathcal{G} \subseteq \mathcal{F}$. The conditional OCE risk measure $\mu_{\phi|\mathcal{G}} : \mathcal{Z} \to L_p(\Omega, \mathcal{G}, \mathbb{P})$ satisfies the following properties:*

*(i) Monotonicity:* $\mu_{\phi|\mathcal{G}}(Z) \leq \mu_{\phi|\mathcal{G}}(W)$*, for every* $Z, W \in \mathcal{Z}$ *such that* $Z \leq W$.

*(ii) Predictable translation equivariance: if* $W \in L_p(\Omega, \mathcal{G}, \mathbb{P})$ *and* $Z \in \mathcal{Z}$*, then* $\mu_{\phi|\mathcal{G}}(W + Z) = W + \mu_{\phi|\mathcal{G}}(Z)$.

*(iii) Convexity: for every* $0 \leq \lambda \leq 1$ *and for every* $Z, W \in \mathcal{Z}$,

$$\mu_{\phi|\mathcal{G}}(\lambda Z + (1-\lambda)W) \leq \lambda\mu_{\phi|\mathcal{G}}(Z) + (1-\lambda)\mu_{\phi|\mathcal{G}}(W). \tag{4.0.3}$$

*Proof.* (i) Take any $Z, W \in \mathcal{Z}$ satisfying $Z \leq W$. Since $\phi(\cdot)$ is monotone, the following inequalities hold for every $Y \in L_p(\Omega, \mathcal{F}, \mathbb{P})$ :

$$\begin{aligned} \mu_{\phi|\mathcal{G}}(Z) &\leq Y + \mathbb{E}\left[\phi(Z-Y)|\mathcal{G}\right] & (4.0.4)\\ &\leq Y + \mathbb{E}\left[\phi(W-Y)|\mathcal{G}\right]. & (4.0.5)\end{aligned}$$

It follows that $\mu_{\phi|\mathcal{G}}(Z) \leq \mu_{\phi|\mathcal{G}}(W)$, which proves (i).

(ii) Let $W \in L_p(\Omega, \mathcal{G}, \mathbb{P})$ and $Z \in \mathcal{Z}$ be given. Then,

$$\begin{aligned} \mu_{\phi|\mathcal{G}}(W+Z) &= \inf_{Y \in L_p(\Omega,\mathcal{G},\mathbb{P})} \left\{Y + \mathbb{E}\left[\phi(Z+W-Y)|\mathcal{G}\right]\right\} & (4.0.6)\\ &= W + \inf_{Y \in L_p(\Omega,\mathcal{G},\mathbb{P})} \left\{Y - W + \mathbb{E}\left[\phi(Z-(Y-W))|\mathcal{G}\right]\right\}. & (4.0.7)\\ &= W + \mu_{\phi|\mathcal{G}}(Z). & (4.0.8)\end{aligned}$$

(iii) Take any $\lambda \in [0,1]$ and $Z, W \in \mathcal{Z}$. Then,

$$\begin{aligned}&\lambda\left(Y_1 + \mathbb{E}\left[\phi\left(Z - Y_1\right)|\mathcal{G}\right]\right) + (1-\lambda)\left(Y_2 + \mathbb{E}\left[\phi\left(W - Y_2\right)|\mathcal{G}\right]\right) = \\ &\quad \lambda Y_1 + (1-\lambda)Y_2 + \mathbb{E}\left[\lambda\phi\left(Z - Y_1\right) + (1-\lambda)\phi\left(W - Y_2\right)|\mathcal{G}\right],\end{aligned}$$

for every $Y_1, Y_2 \in L_p(\Omega, \mathcal{G}, \mathbb{P})$. Since $\phi$ is convex, it follows that

$$\phi\left(\lambda Z + (1-\lambda)W - (\lambda Y_1 + (1-\lambda)Y_2)\right) \leq \lambda\phi\left(Z - Y_1\right) + (1-\lambda)\phi\left(W - Y_2\right).$$

By the monotonicity of $\mathbb{E}\left[\cdot|\mathcal{G}\right]$, we obtain that:

$$\begin{aligned}&\lambda\left(Y_1 + \mathbb{E}\left[\phi\left(Z - Y_1\right)|\mathcal{G}\right]\right) + (1-\lambda)\left(Y_2 + \mathbb{E}\left[\phi\left(W - Y_2\right)|\mathcal{G}\right]\right)\\ &\quad \geq \lambda Y_1 + (1-\lambda)Y_2 + \mathbb{E}\left[\phi\left(\lambda Z + (1-\lambda)W - (\lambda Y_1 + (1-\lambda)Y_2)\right)|\mathcal{G}\right]\\ &\quad\quad \geq \mu_{\phi|\mathcal{G}}(\lambda Z + (1-\lambda)W).\end{aligned}$$

Since $Y_1$ and $Y_2$ are arbitrary, it follows that $\mu_{\phi|\mathcal{G}}(\lambda Z + (1-\lambda)W) \leq \lambda\mu_{\phi|\mathcal{G}}(Z) + (1-\lambda)\mu_{\phi|\mathcal{G}}(W)$. □

The nested formulation of risk neutral multistage stochastic programming problems (see (2.1.58)) makes use of the conditional expectation as the optimization criterion at each stage $t = 1, \ldots, T$ of the decision process. At each stage $t$, new information $\xi_t$ becomes available to the optimizer before he makes the $t^{th}$-stage decision $x_t = x_t(\xi_{[t]})$[1]. As before, assume that the optimization problem has random data $\xi = (\xi_1, \xi_2, \ldots, \xi_T)$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $T \geq 3$ is an integer. The flow of information is modeled through the filtration $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_T$ generated by the history process $\xi_{[t]}$ up to time $t$, for $t = 1, \ldots, T$, i.e.,

$$\mathcal{F}_t := \sigma(\xi_1, \ldots, \xi_t), \text{ for } t = 1, \ldots, T. \tag{4.0.9}$$

We assume that $\mathcal{F}_1 = \{\emptyset, \Omega\}$, i.e., $\xi_1$ is deterministic, and that $\mathcal{F}_T = \mathcal{F}$. Given $\phi_t \in \Phi$, for $t = 2, \ldots, T$, we consider the conditional OCE risk measures

$$\begin{array}{rccl} \mu_{\phi_t|\mathcal{F}_{t-1}}: & \mathcal{Z}_t & \to & \mathcal{Z}_{t-1} \\ & Z & \mapsto & \mu_{\phi_t|\mathcal{F}_{t-1}}(Z) = \inf\limits_{Y \in \mathcal{Z}_{t-1}} \left\{ Y + \mathbb{E}\phi_t(Z - Y) \right\}, \end{array} \tag{4.0.10}$$

where $\mathcal{Z}_t = L_p(\Omega, \mathcal{F}_t, \mathbb{P})$, for every $t = 1, \ldots, T$; $p \in [1, \infty)$; and $\phi_t \in \Phi$ is such that $|\phi_t(Z)|^p$ is integrable, for every $Z \in \mathcal{Z}_t$. When we present our results, we always suppose that $\phi_t$ is Lipschitz continuous, for every $t = 2, \ldots, T$. So, we can always take $p = 1$ above (see also Remark 4.0.19), since $\phi_t(Z)$ is integebrable, for every $Z \in \mathcal{Z}_t$. We also denote the conditional OCE risk measures above as $\mu_{\phi_{t+1}|\xi_{[t]}}$, for $t = 1, \ldots, T-1$, since $\mathcal{F}_t$ is simply the $\sigma$-algebra generated by $\xi_{[t]}$. The general risk averse $T$-stage stochastic programming problem with nested OCE risk measures is formulated as

$$\min_{x_1 \in X_1} \left\{ F_1(x_1) + \mu_{\phi_2|\xi_1} \left( \inf_{x_2 \in X_2(x_1, \xi_2)} F_2(x_2, \xi_2) \quad + \mu_{\phi_3|\xi_{[2]}} \left( \ldots \right.\right.\right.$$
$$\left.\left.\left. + \mu_{\phi_T|\xi_{[T-1]}} \left( \inf_{x_T \in X_T(x_{T-1}, \xi_T)} F_T(x_T, \xi_T) \right) \ldots \right) \right\}, \tag{4.0.11}$$

where $x_t \in \mathbb{R}^{n_t}, t = 1, ..., T$, are the decisions variables, $F_t : \mathbb{R}^{n_t} \times \mathbb{R}^{d_t} \to \mathbb{R}$, $t = 2, \ldots, T$, are Carathéodory functions, and $X_t : \mathbb{R}^{n_{t-1}} \times \mathbb{R}^{d_t} \rightrightarrows \mathbb{R}^{n_t}$, $t = 2, ..., T$, are closed-valued measurable multifunctions. We assume that the function $F_1 : \mathbb{R}^{n_1} \to \mathbb{R}$ is continuous, and $X_1 \subseteq \mathbb{R}^{n_1}$ is a nonempty closed set. Unless otherwise stated, all these features are automatically assumed in the remainder of this chapter.

Beginning in the last stage $t = T$, we can write the dynamic programming equations

$$Q_T(x_{T-1}, \xi_T) = \inf_{x_T \in X_T(x_{T-1}, \xi_T)} F_T(x_T, \xi_T). \tag{4.0.12}$$

Since $F_T : \mathbb{R}^{n_T} \times \mathbb{R}^{d_T} \to \mathbb{R}$ is a Carathéodory function and $X_T(x_{T-1}, \cdot)$ is a closed-valued measurable multifunction, we have by Corollary 2.6.17 (see also Proposition 2.6.15) that $Q_T(x_{T-1}, \cdot)$ is measurable, for every $x_{T-1}$. Let us assume that $Q_T(x_{T-1}, \cdot)$ is also integrable, for every $x_{T-1}$. Then, $Q_T(x_{T-1}, \cdot) \in \mathcal{Z}_T$ and

$$\mathcal{Q}_T(x_{t-1}, \xi_{[t-1]}) := \mu_{\phi_T|\xi_{[T-1]}}(Q_T(x_{T-1}, \xi_T)) \in \mathcal{Z}_{T-1}, \tag{4.0.13}$$

[1]At stage $t = 1$, the optimizer already knows the value of $\xi_1$, since it is deterministic.

supposing, for example, that $L(\phi_T) < \infty$. Having considered the function $\mathcal{Q}_{t+1}(x_t, \xi_{[t+1]})$ for some $t = T-1, \ldots, 2$, we write the dynamic programming equation

$$Q_t(x_{t-1}, \xi_{[t]}) = \inf_{x_t \in X_t(x_{t-1}, \xi_t)} \left\{ F_t(x_t, \xi_t) + \mathcal{Q}_{t+1}(x_t, \xi_{[t]}) \right\} \tag{4.0.14}$$

that we assume belongs to $\mathcal{Z}_t$. Then, we consider the function

$$\mathcal{Q}_t(x_{t-1}, \xi_{[t-1]}) := \mu_{\phi_t|\xi_{[t-1]}} \left( Q_t(x_{t-1}, \xi_{[t]}) \right) \in \mathcal{Z}_{t-1}, \tag{4.0.15}$$

assuming, for example, that $L(\phi_t) < \infty$. At the first stage one solves

$$\min_{x_1 \in X_1} \left\{ F_1(x_1) + \mathcal{Q}_2(x_1, \xi_1) \right\}. \tag{4.0.16}$$

Note that $\omega \in \Omega \mapsto \mathcal{Q}_2(x_1, \xi_1(\omega))$ is constant, since $\xi_1$ is deterministic.

One can consider Monte Carlo sampling-based approaches to approximate problem (4.0.11) by a problem driven by a finite state random data $\hat{\xi} = \left(\hat{\xi}_1, \hat{\xi}_2, \ldots, \hat{\xi}_T\right)$ (the empirical data generated by the sampling scheme). In the sequel we derive sample complexity results for $T$-stage problems like (4.0.11) considering suitable regularity conditions. Before proceeding, it is worth making two points about the problems studied here.

First, let us recall that the nested structure considered in (4.0.11) is in some sense not very restrictive. Here we make a very short presentation about this topic. It is possible to develop a general theory of dynamic risk measures $\{\rho_{t,T} : \mathcal{Z}_{t,T} \to \mathcal{Z}_t\}_{t=1}^T$ that are used for in a sequential decision making process for evaluating a sequence of random outcomes $Z_1 \in \mathcal{Z}_1, Z_2 \in \mathcal{Z}_2, \ldots, Z_T \in \mathcal{Z}_T$ (see, for instance, [62] and the references therein), where $\mathcal{Z}_{t,T} = \mathcal{Z}_t \times \cdots \times \mathcal{Z}_T$, for $t = 1, \ldots, T$. Considering the dynamic risk measures, a multistage stochastic programming problem is formulated as:

$$\begin{array}{rl} \min & \rho_{1,T}\left(F_1(x_1), F_2(x_2(\xi_{[2]}), \xi_2), \ldots, F_T(x_T(\xi_{[2]}), \xi_T)\right) \\ s.t. & x_t = x_t(\xi_{[t]}), \text{ for } t = 1, \ldots, T \\ & x_1 \in X_1, \text{ and } x_t(\xi_{[t]}) \in X_t(x_{t-1}(\xi_{t-1}), \xi_t), \text{ for } t = 2, \ldots, T. \end{array} \tag{4.0.17}$$

A key concept in the development of the theory of dynamic risk measures is the notion of time consistency (see [73, Definition 6.76]). Regarding this notion, it is possible to show that (see [62, Theorem 1] or [73, Theorem 6.78]) if $\{\rho_{t,T} : \mathcal{Z}_{t,T} \to \mathcal{Z}_t\}_{t=1}^T$ are time consistent dynamic risk measures satisfying some side conditions, such as the *predictable translation equivariance* condition and the *normalization* condition $\rho_{t,T}(0, \ldots, 0) = 0$, then

$$\rho_{t,T}(Z_t, \ldots, Z_T) = Z_t + \rho_t\left(Z_{t+1} + \rho_{t+1}\left(Z_{t+2} + \rho_{t+2}\left(\ldots + \rho_{T-1}(Z_T)\ldots\right)\right)\right), \tag{4.0.18}$$

where $\rho_s : \mathcal{Z}_{s+1} \to \mathcal{Z}_s$ is a one-step conditional risk measures, for $s = 1, \ldots, T-1$.

Second, many publications dealing with risk averse multistage stochastic programming problems like [19, 39, 51, 70, 72, 76] consider problems having a nested structure like (4.0.11). In these references, a one-step mean-risk model using the Average Value-at-Risk risk measure is usually considered:

$$\rho_{t|\xi_{[t-1]}}(Z) := (1-\lambda_t)\mathbb{E}\left[Z|\xi_{[t-1]}\right] + \lambda_t \,\mathrm{AV@R}_{1-\alpha_t}\left(Z|\xi_{[t-1]}\right),$$

where $\lambda_t \in [0,1]$ and $\alpha_t \in (0,1)$ are chosen parameters, for $t = 2, \ldots, T$. Moreover, in practical applications one usually considers risk averse multistage linear programming problems, i.e., problems satisfying: $F_t(x_t, \xi_t) = \langle c_t, x_t\rangle$ and $X_t(x_{t-1}, \xi_t) = \{x_t \in \mathbb{R}^{n_t} :\ A_t x_t + B_t x_{t-1} = b_t, x_t \geq 0\}$, where $\xi_t = (c_t, A_t, B_t, b_t)$ for each $t = 2, \ldots,$ $T$. Naturally, one also assumes that $F_1(x_1) = \langle c_1, x_1\rangle$ and $X_1 = \{x_1 \in \mathbb{R}^{n_1} : A_1 x_1 = b_1, x_1 \geq 0\}$. Thus, the framework considered in (4.0.11) encompasses an important class of risk averse stochastic programming problems considered by the stochastic programming community.

Akin to risk neutral multistage stochastic programming problems, here we also assume that the random data $(\xi_1, \ldots, \xi_T)$ is stagewise independent. In the next proposition we show that under this hypothesis the one-step conditional OCE risk measures $\mu_{\phi_t|\xi_{[t-1]}}(\cdot)$ boil down to the regular OCE risk measure $\mu_{\phi_t}(\cdot)$. This is an expected result like the one observed for the conditional expected value operator. The following lemma will be useful.

**Lemma 4.0.21.** *Let $Y$ and $Z$ be independent random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\varphi : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ be a function satisfying $\mathbb{E}\,|\varphi(Y,Z)| < \infty$ and let $\Upsilon(y) = \mathbb{E}\varphi(y, Z)$. Then, the following identity is satisfied*

$$\mathbb{E}\left[\varphi(Y,Z)|Y\right] = \Upsilon(Y). \tag{4.0.19}$$

*Proof.* See [22, Example 5.1.5]. □

Lemma 4.0.21 can be applied in a straightforward way for proving the result regarding the equality between the conditional OCE risk measure and the regular OCE risk measure.

**Proposition 4.0.22.** *Let $g : \mathbb{R}^{d_t} \to \mathbb{R}$ be a Borel-measurable function such that $Z = g(\xi_t)$ is an integrable random variable, where $2 \leq t \leq T$. Take any $\phi \in \Phi$ such that $\phi(Z)$ is integrable. If $\xi_t$ is independent of $\mathcal{F}_{t-1} = \sigma(\xi_1, \ldots, \xi_{t-1})$, then*

$$\mu_{\phi|\xi_{[t-1]}}(Z) = \mu_\phi(Z). \tag{4.0.20}$$

*Proof.* First note that for every $s \in \mathbb{R}$, we have that $Y := s \cdot \mathbb{1}_\Omega(\cdot) \in \mathcal{Z}_{t-1}$. So, for every $Z \in \mathcal{Z}_t$, we have that

$$s + \mathbb{E}\phi(Z - s) = Y + \mathbb{E}\left[\phi\left(Z - Y\right)|\mathcal{F}_{t-1}\right]. \tag{4.0.21}$$

It follows that $\mu_{\phi|\xi_{[t-1]}}(Z) \leq \mu_\phi(Z)$.

Now, we prove that the converse inequality also holds under the hypotheses of the present proposition. Take any $Y \in \mathcal{Z}_{t-1}$. Thus, there exists a Borel measurable function $h$ such that $Y = h(\xi_1, \ldots, \xi_{t-1})$. Define $\varphi(\xi_{[t-1]}, \xi_t) := h(\xi_{[t-1]}) + \phi\left(g(\xi_t) - h(\xi_{[t-1]})\right)$. Note that

$$
\begin{aligned}
Y(\omega) + \mathbb{E}\left[\phi(Z-Y)|\xi_{[t-1]}\right](\omega) &= \mathbb{E}\left[Y + \phi(Z-Y)|\xi_{[t-1]}\right](\omega) && (4.0.22)\\
&= \mathbb{E}\left[\varphi(\xi_{[t-1]}, \xi_t)|\xi_{[t-1]}\right](\omega) && (4.0.23)\\
&= \Upsilon(\xi_{[t-1]}(\omega)) && (4.0.24)\\
&= Y(\omega) + \mathbb{E}\phi(Z - Y(\omega)) && (4.0.25)\\
&\geq \inf_{s\in\mathbb{R}} \{s + \mathbb{E}\phi(Z-s)\} && (4.0.26)\\
&= \mu_\phi(Z), && (4.0.27)
\end{aligned}
$$

where (4.0.24) follows from Lemma 4.0.21 with

$$
\Upsilon(e_1, \ldots, e_{t-1}) = \mathbb{E}\varphi(e_1, \ldots, e_{t-1}, \xi_t) = h(e_1, \ldots, e_{t-1}) + \mathbb{E}\phi\left(Z - h(e_1, \ldots, e_{t-1})\right).
$$

This concludes the proof of the proposition. $\square$

Assuming the stagewise independence condition the $T$-stage stochastic programming problem becomes

$$
\begin{aligned}
\min_{x_1\in X_1} \Bigg\{ F_1(x_1) + \mu_{\phi_2}\Bigg( &\inf_{x_2\in X_2(x_1,\xi_2)} F_2(x_2,\xi_2) \quad + \mu_{\phi_3}\Bigg( \ldots \\
&+\mu_{\phi_T}\Bigg(\inf_{x_T\in X_T(x_{T-1},\xi_T)} F_T(x_T,\xi_T)\Bigg) \ldots \Bigg)\Bigg\}.
\end{aligned} \tag{4.0.28}
$$

Now, let us see how the stagewise independence condition affects the cost-to-go functions. Beginning in the last-stage, the dynamic programming equation under the stage independence hypothesis becomes

$$
\begin{aligned}
\mathcal{Q}_T(x_{T-1}, \xi_{[T-1]}) &:= \mu_{\phi_T|\xi_{[T-1]}}\left(Q_T(x_{T-1}, \xi_T)\right)\\
&= \mu_{\phi_T}\left(Q_T(x_{T-1}, \xi_T)\right).
\end{aligned} \tag{4.0.29}
$$

This means that $\mathcal{Q}_T(x_{T-1}, \xi_{[T-1]}) = \mathcal{Q}_T(x_{T-1})$, i.e., it does not depend on $\xi_{[T-1]}$. So, $Q_{T-1}(x_{T-2}, \xi_{[T-1]})$ does not depend on the entire history process $\xi_{[T-1]}$ up to stage $T-1$, but just on $\xi_{T-1}$. Continuing backward in stages, we obtain that

$$
Q_t(x_{t-1}, \xi_t) = \inf_{x_t\in X_t(x_{t-1},\xi_t)} \{F_t(x_t, \xi_t) + \mathcal{Q}_{t+1}(x_t)\}, \text{ and} \tag{4.0.30}
$$

$$
\mathcal{Q}_t(x_{t-1}) := \mu_{\phi_t}\left(Q_t(x_{t-1}, \xi_t)\right), \tag{4.0.31}
$$

for every $t = T-1, \ldots, 2$.

The optimal value of problem (4.0.28) is

$$v^* := \inf_{x_1 \in X_1} \{v(x_1) := F_1(x_1) + \mathcal{Q}_2(x_1)\} \tag{4.0.32}$$

and its set of $\epsilon$-solutions is given by

$$S^\epsilon := \{x_1 \in X_1 : v(x_1) \le v^* + \epsilon\}, \tag{4.0.33}$$

where $\epsilon \ge 0$.

Similarly to the analysis in Section 2.1.2, here we approximate the random data constructing a scenario tree via the *identical conditional sampling* scheme. As before, $N_t \in \mathbb{N}$ is the number of samples realizations in the $t^{th}$-stage, for $t = 2, \ldots, T$. Given any sample realization $\left\{\xi_t^j : j = 1, \ldots, N_t, t = 2, \ldots, T\right\}$, the associated stochastic process $\hat{\xi} = \left(\hat{\xi}_1, \ldots, \hat{\xi}_T\right)$ is stagewise independent (see Proposition 2.1.12). The empirical or the "SAA" problem is just

$$\begin{aligned}\min_{x_1 \in X_1} \Bigg\{ F_1(x_1) + \hat{\mu}_{\phi_2} \Bigg( \inf_{x_2 \in X_2(x_1, \hat{\xi}_2)} F_2(x_2, \hat{\xi}_2) \quad + \hat{\mu}_{\phi_3} \Bigg( \ldots \\ + \hat{\mu}_{\phi_T} \Bigg( \inf_{x_T \in X_T(x_{T-1}, \hat{\xi}_T)} F_T(x_T, \hat{\xi}_T) \Bigg) \ldots \Bigg) \Bigg\}.\end{aligned} \tag{4.0.34}$$

Akin to Corollary 2.1.14, we obtain the following formulas for the empirical cost-to-go functions in the multistage setting with OCE risk measures

$$\hat{Q}_t\left(x_{t-1}, \xi_t^j\right) = \inf_{x_t \in X_t\left(x_{t-1}, \xi_t^j\right)} \left\{F_t\left(x_t, \xi_t^j\right) + \hat{\mathcal{Q}}_{t+1}(x_t)\right\} \tag{4.0.35}$$

$$\hat{\mathcal{Q}}_t(x_{t-1}) = \hat{\mu}_{\phi_t}\left(\hat{Q}_t\left(x_{t-1}, \hat{\xi}_t\right)\right) \tag{4.0.36}$$

$$= \inf_{s \in \mathbb{R}} \left\{ s + \frac{1}{N_t} \sum_{j=1}^{N_t} \phi_t\left(\hat{Q}_t\left(x_{t-1}, \xi_t^j\right) - s\right)\right\}, \tag{4.0.37}$$

for $1 \le j \le N_t$ and $t = 2, \ldots, T$. As usual, we consider the boundary condition $\hat{\mathcal{Q}}_{T+1}(x_T) = 0$, for every $x_T \in \mathbb{R}^{n_T}$.

The optimal value of problem (4.0.34) is

$$\hat{v}^*_{N_2,\ldots,N_T} := \inf_{x_1 \in X_1} \left\{\hat{v}_{N_2,\ldots,N_T}(x_1) := F_1(x_1) + \hat{\mathcal{Q}}_2(x_1)\right\} \tag{4.0.38}$$

and its set of $\epsilon$-solutions is given by

$$\hat{S}^\epsilon_{N_2,\ldots,N_T} := \{x_1 \in X_1 : \hat{v}_{N_2,\ldots,N_T}(x_1) \le \hat{v}^*_{N_2,\ldots,N_T} + \epsilon\}, \tag{4.0.39}$$

where $\epsilon \ge 0$. In the sequel we derive sample estimates $N_2, \ldots, N_T$ in order to guarantee that

$$\mathbb{P}\left[\hat{S}^\delta_{N_2,\ldots,N_T} \subseteq S^\epsilon\right] \ge 1 - C(\epsilon, \delta) \exp\left\{-\beta(\epsilon, \delta) N\right\}, \text{ for every } N \in \mathbb{N}, \tag{4.0.40}$$

where $C(\epsilon, \delta)$ and $\beta(\epsilon, \delta)$ are positive constants that depend on the sample complexity parameters $0 \leq \delta < \epsilon$ and on the problem data.

As before, we consider the following notation to be used in the sequel:

$$\begin{aligned} \mathcal{X}_0 &:= \{0\} \subseteq \mathbb{R}, \\ X_1(x_0, \xi_1) &:= X_1, \ \forall x_0 \in \mathcal{X}_0, \\ \mathcal{Q}_{T+1}(x_T) &:= 0, \ \forall x_T \in \mathbb{R}^{n_{T+1}}. \end{aligned}$$

Below we enumerate the same regularity conditions used in Section 2.1.2 for deriving the sample complexity estimates for a risk neutral $T$-stage stochastic programming problem:

(M0) The random data $\xi_1, \xi_2, \ldots, \xi_T$ is stagewise independent.

(M1) The family of random vectors $\{\xi_t^j : j \in \mathbb{N}, t = 2, \ldots, T\}$ is independent and satisfies $\xi_t^j \stackrel{d}{\sim} \xi_t$, for all $j \in \mathbb{N}$, and $t = 2, \ldots, T$.

For each $t = 1, \ldots, T-1$:

(Mt.1) There exist a compact set $\mathcal{X}_t$ with finite diameter $D_t$ such that $X_t(x_{t-1}, \xi_t) \subseteq \mathcal{X}_t$, for every $x_{t-1} \in \mathcal{X}_{t-1}$ and $\xi_t \in \operatorname{supp}(\xi_t)$.

(Mt.2) $\mathbb{E}Q_{t+1}(x_t, \xi_{t+1})$ is finite, for every $x_t \in \mathcal{X}_t$.

(Mt.3) There exists a finite constant $\sigma_t > 0$ such that for any $x \in \mathcal{X}_t$, the following inequality holds

$$M_{t,x}(s) := \mathbb{E}\left[\exp\left(s(Q_{t+1}(x, \xi_{t+1}) - \mathbb{E}Q_{t+1}(x, \xi_{t+1})\right)\right] \leq \exp\left(\sigma_t^2 s^2/2\right), \ \forall s \in \mathbb{R}. \tag{4.0.41}$$

(Mt.4) There exists a measurable function $\chi_t : \operatorname{supp}(\xi_{t+1}) \to \mathbb{R}_+$ whose moment generating function $M_{\chi_t}(s)$ is finite, for $s$ in a neighborhood of zero, such that

$$\left|Q_{t+1}(x_t', \xi_{t+1}) - Q_{t+1}(x_t, \xi_{t+1})\right| \leq \chi_t(\xi_{t+1}) \left|\left|x_t' - x_t\right|\right| \tag{4.0.42}$$

holds, for all $x_t', x_t \in \mathcal{X}_t$ and $\xi_{t+1} \in E_{t+1} \subseteq \operatorname{supp} \xi_{t+1}$, where $\mathbb{P}\left[\xi_{t+1} \in E_{t+1}\right] = 1$.

(Mt.5) W.p.1 $\xi_{t+1}$ the multifunction $X_{t+1}(\cdot, \xi_{t+1})$ restricted to $\mathcal{X}_t$ is continuous.

Remarks regarding these regularity conditions were already made in Section 2.1.2 and we do not repeat them here. As we have done in the risk neutral multistage setting we first show that under these regularity conditions problems (4.0.28) and (4.0.34) are solvable. As before, whenever we assume conditions (Mt.4), for $t = 1, \ldots, T-1$, we denote the expected value of $\chi_t(\xi_{t+1})$ as

$$0 \leq M_t := \mathbb{E}\chi_t(\xi_{t+1}) < \infty. \tag{4.0.43}$$

We are ready to prove the following proposition.

**Proposition 4.0.23.** *Consider a general $T$-stage stochastic programming problem such as (4.0.11), where $T \geq 3$ is an arbitrary integer. Assume that $\phi_t \in \Phi$ satisfies $L(\phi_t) < \infty$, for every $t = 2, \ldots, T$. The following assertions hold:*

*(a) If the problem satisfies the regularity conditions (M0), (Mt.1), (Mt.2), and (Mt.4), for $t = 1, \ldots, T-1$, then $Q_{t+1}(\cdot, \xi_{t+1})$ is a Lipschitz continuous function on $\mathcal{X}_t$ w.p.1 $\xi_{t+1}$, for $t = 1, \ldots, T-1$. It also follows that $\mathcal{X}_t \subseteq \operatorname{dom} X_{t+1}(\cdot, \xi_{t+1})$ w.p.1 $\xi_{t+1}$ and $\mathcal{Q}_{t+1}(\cdot)$ is $(M_t L(\phi_{t+1}))$-Lipschitz continuous on $\mathcal{X}_t$, for $t = 1, \ldots, T-1$. In particular, we conclude that the first stage objective function*

$$v(x_1) = F_1(x_1) + \mathcal{Q}_2(x_1)$$

*of the true problem restricted to $x_1 \in \mathcal{X}_1$ is finite-valued and continuous and the set of first stage optimal solutions $S$ is nonempty.*

*(b) Consider as given the sample sizes $N_2, \ldots, N_T \in \mathbb{N}$. If the problem satisfies the regularity conditions (M0), (M1), (Mt.1), (Mt.4) and (Mt.5), for $t = 1, \ldots, T-1$, and the SAA scenario tree is constructed using the* identical conditional sampling *scheme, then the SAA objective function $\hat{v}_{N_2,\ldots,N_T}(x_1)$ restricted to the set $\mathcal{X}_1$ is finite-valued and continuous w.p.1. In particular, $\mathbb{P}\left[\hat{S}_{N_2,\ldots,N_T} \neq \emptyset\right] = 1$.*

*Proof.* The proof of this proposition is similar to the one of Proposition 3.3.12. Let us begin with item (a). Condition (Mt.4) implies that, for every $\xi_{t+1} \in E_{t+1}$, $Q_{t+1}(\cdot, \xi_{t+1})$ is $\chi_t(\xi_{t+1})$-Lipschitz continuous on $\mathcal{X}_t$, where $E_{t+1} \subseteq \operatorname{supp} \xi_{t+1}$ satisfies

$$\mathbb{P}\left[\xi_{t+1} \in E_{t+1}\right] = 1,$$

for $t = 1, \ldots, T-1$. In particular, we have that $Q_{t+1}(\cdot, \xi_{t+1})$ is a finite-valued function on $\mathcal{X}_t$, for every $\xi_{t+1} \in E_{t+1}$. Since

$$Q_{t+1}(x_t, \xi_{t+1}) = \inf_{x_{t+1} \in X_{t+1}(x_t, \xi_{t+1})} \left\{F_{t+1}(x_{t+1}, \xi_{t+1}) + \mathcal{Q}_{t+2}(x_{t+1})\right\}, \tag{4.0.44}$$

it follows that $X_{t+1}(x_t, \xi_{t+1}) \neq \emptyset$, for all $x_t \in \mathcal{X}_t$ and $\xi_{t+1} \in E_{t+1}$, i.e. $\operatorname{dom} X_{t+1}(\cdot, \xi_{t+1}) \supseteq \mathcal{X}_t$, for all $t = 1, \ldots, T-1$. Assuming conditions (Mt.2), for $t = 1, \ldots, T-1$, it follows that

$$\begin{aligned} \left\|Q_{t+1}(x_t', \xi_{t+1}) - Q_{t+1}(x_t, \xi_{t+1})\right\|_1 &\leq \mathbb{E}\chi_t(\xi_{t+1}) \left\|x_t' - x_t\right\| & (4.0.45)\\ &= M_t \left\|x_t' - x_t\right\|, & (4.0.46) \end{aligned}$$

By Proposition 3.1.9 we have that $\mu_{\phi_{t+1}} : \mathcal{Z} \to \mathbb{R}$ is $L(\phi_{t+1})$-Lipschitz continuous, since $L(\phi_{t+1}) < \infty$. So,

$$\begin{aligned} \left|\mathcal{Q}_{t+1}(x_t') - \mathcal{Q}_{t+1}(x_t)\right| &= \left|\mu_{\phi_{t+1}}\left(Q_{t+1}(x_t', \xi_{t+1})\right) - \mu_{\phi_{t+1}}\left(Q_{t+1}(x_t, \xi_{t+1})\right)\right| & (4.0.47)\\ &\leq L(\phi_{t+1}) \left\|Q_{t+1}(x_t', \xi_{t+1}) - Q_{t+1}(x_t, \xi_{t+1})\right\|_1 & (4.0.48)\\ &\leq L(\phi_{t+1}) M_t \left\|x_t' - x_t\right\|, & (4.0.49) \end{aligned}$$

for every $x_t', x_t \in \mathcal{X}_t$. The first stage objective function of the true problem is just $v(x_1) = F_1(x_1) + \mathcal{Q}_2(x_1)$, where $F_1 : \mathbb{R}^n \to \mathbb{R}$ is a finite-valued continuous function. It follows that the restriction of $v$ to the compact set $\mathcal{X}_1 \supseteq X_1$ is continuous and that $S = \operatorname{argmin}_{x_1 \in X_1} v(x_1) \neq \emptyset$, since $X_1$ is nonempty compact. This completes the proof of item (a).

Now, we prove item (b). Condition (Mt.5) says that there exists $F_{t+1} \subseteq \operatorname{supp} \xi_{t+1}$ satisfying $\mathbb{P}\left[\xi_{t+1} \in F_{t+1}\right] = 1$ such that $X_{t+1}(\cdot, \xi_{t+1}) : \mathcal{X}_t \rightrightarrows \mathcal{X}_{t+1}$ is a continuous multifunction, for every $\xi_{t+1} \in F_{t+1}$ and for every $t = 1, \ldots, T-1$. Since conditions (M1) and (Mt.4) also hold true, for $t = 1, \ldots, T-1$, the following event has probability 1

$$\mathcal{E} := \bigcap_{t=1}^{T-1} \bigcap_{j \in \mathbb{N}} \left[\xi_{t+1}^j \in E_{t+1} \cap F_{t+1}\right]. \tag{4.0.50}$$

Take any sample sizes $N_2, \ldots, N_T \in \mathbb{N}$. We have that

$$\mathcal{E}_{N_2,\ldots,N_T} := \bigcap_{t=1}^{T-1} \bigcap_{j=1}^{N_{t+1}} \left[\xi_{t+1}^j \in E_{t+1} \cap F_{t+1}\right] \supseteq \mathcal{E}, \tag{4.0.51}$$

therefore $\mathbb{P}\left(\mathcal{E}_{N_2,\ldots,N_T}\right) = 1$.

Now, we show that whenever the event $\mathcal{E}_{N_2,\ldots,N_T}$ happens, every function

$$\begin{aligned}
\hat{\mathcal{Q}}_{t+1}(x_t) &= \hat{\mu}_{\phi_{t+1}}\left(\hat{Q}_{t+1}\left(x_t, \hat{\xi}_{t+1}\right)\right) \\
&= \inf_{s \in \mathbb{R}} \left\{ s + \frac{1}{N_{t+1}} \sum_{j=1}^{N_{t+1}} \phi_{t+1}\left(\hat{Q}_{t+1}(x_t, \xi_{t+1}^j) - s\right) \right\}, \ \forall x_t \in \mathcal{X}_t,
\end{aligned}$$

is finite-valued and continuous on $\mathcal{X}_t$, for $t = 1, \ldots, T-1$, where

$$\hat{Q}_{t+1}(x_t, \xi_{t+1}^j) = \inf_{x_{t+1} \in X_{t+1}(x_t, \xi_{t+1}^j)} \left\{ F_{t+1}(x_{t+1}, \xi_{t+1}) + \hat{\mathcal{Q}}_{t+2}(x_{t+1}) \right\} \tag{4.0.52}$$

are the empirical cost-to-go functions, for $t = 1, \ldots, T-1$. As usual we set $\hat{\mathcal{Q}}_{T+1}(x_T) := 0$, for every $x_T \in \mathcal{X}_T$, for uniformity of notation. For proving the result we show that if $\hat{\mathcal{Q}}_{t+1} : \mathcal{X}_t \to \mathbb{R}$ is finite-valued and continuous, then $\hat{\mathcal{Q}}_t : \mathcal{X}_{t-1} \to \mathbb{R}$ is also finite-valued and continuous, for $t = T, \ldots, 2$. We also verify that $\hat{\mathcal{Q}}_T(\cdot) : \mathcal{X}_{T-1} \to \mathbb{R}$ is finite-valued and continuous (base case in order to apply the induction step). Note that

$$\hat{Q}_T(x_{T-1}, \xi_T) = Q_T(x_{T-1}, \xi_T), \tag{4.0.53}$$

for every $x_{T-1} \in \mathcal{X}_{T-1}$ and $\xi_T \in \operatorname{supp} \xi_T$.

Whenever the event $\mathcal{E}_{N_2,\ldots,N_T}$ occurs, we have in particular that $\xi_T^j \in E_T$, for every $j = 1, \ldots, N_T$. So, it follows from item (a) that

$$x_{T-1} \in \mathcal{X}_{T-1} \mapsto Q_T(x_{T-1}, \xi_T^j) \tag{4.0.54}$$

is $\chi_{T-1}(\xi_T^j)$-Lipschitz continuous on $\mathcal{X}_{T-1}$, for every $j = 1, \ldots, N_T$. Therefore, the mapping

$$x_{T-1} \in \mathcal{X}_{T-1} \mapsto Q_T(x_{T-1}, \hat{\xi}_T) \in L_1\left(\hat{\mathcal{S}}, \mathcal{P}(\hat{\mathcal{S}}), \hat{\mathbb{P}}\right) \tag{4.0.55}$$

satisfies

$$\begin{aligned}\left\|\hat{Q}_T(x'_{T-1}, \hat{\xi}_T) - \hat{Q}_T(x_{T-1}, \hat{\xi}_T)\right\|_{L_1\left(\hat{\mathcal{S}}, \mathcal{P}(\hat{\mathcal{S}}), \hat{\mathbb{P}}\right)} &= \frac{1}{N_T}\sum_{j=1}^{N_T}\left|Q_T(x'_{T-1}, \xi_T^j) - Q_T(x_{T-1}, \xi_T^j)\right| \\ &\leq \left(\frac{1}{N_T}\sum_{j=1}^{N_T}\chi_{T-1}(\xi_T^j)\right)\left\|x'_{T-1} - x_{T-1}\right\|,\end{aligned}$$

for every $x'_{T-1}, x_{T-1} \in \mathcal{X}_{T-1}$. In particular, the mapping (4.0.55) is continuous. Since $L(\phi_T) < \infty$, we also have from Proposition 3.1.9 that $\hat{\mu}_{\phi_T} : L_1\left(\hat{\mathcal{S}}, \mathcal{P}(\hat{\mathcal{S}}), \hat{\mathbb{P}}\right) \to \mathbb{R}$ is $L(\phi_T)$-Lipschitz continuous. This proves that

$$x_{T-1} \in \mathcal{X}_{T-1} \mapsto \hat{\mathcal{Q}}_T(x_{T-1}), \tag{4.0.56}$$

is continuous, which is the base case.

Now we prove the induction step. Assume that $\hat{\mathcal{Q}}_{t+1} : \mathcal{X}_t \to \mathbb{R}$ is finite-valued and continuous for some $t + 1 \leq T$[2]. We claim that

$$\hat{Q}_t(x_{t-1}, \xi_t^j) = \inf_{x_t \in X_t(x_{t-1}, \xi_t^j)}\left\{F_t(x_t, \xi_t^j) + \hat{\mathcal{Q}}_{t+1}(x_t)\right\}, \forall x_{t-1} \in \mathcal{X}_{t-1}, \tag{4.0.57}$$

is finite-valued and continuous, for every $j = 1, \ldots, N_t$, whenever the event $\mathcal{E}_{N_2,\ldots,N_T}$ happens. This fact follows from the BMT (see Proposition 2.6.4) exactly in the same way as considered in Proposition 2.1.15. Since

$$\left\|\hat{Q}_t(x'_{t-1}, \hat{\xi}_t) - \hat{Q}_t(x_{t-1}, \hat{\xi}_t)\right\|_{L_1\left(\hat{\mathcal{S}}, \mathcal{P}(\hat{\mathcal{S}}), \hat{\mathbb{P}}\right)} = \frac{1}{N_t}\sum_{j=1}^{N_t}\left|\hat{Q}_t(x'_{t-1}, \xi_t^j) - \hat{Q}_t(x_{t-1}, \xi_t^j)\right|,$$

it follows that $x_{t-1} \in \mathcal{X}_{t-1} \mapsto \hat{Q}_t(x_{t-1}, \hat{\xi}_t) \in L_1\left(\hat{\mathcal{S}}, \mathcal{P}(\hat{\mathcal{S}}), \hat{\mathbb{P}}\right)$. Since $L(\phi_t) < \infty$, we conclude that the mapping

$$x_{t-1} \in \mathcal{X}_{t-1} \mapsto \hat{\mathcal{Q}}_t(x_{t-1}) = \hat{\mu}_{\phi_t}\left(\hat{Q}_t(x_{t-1}, \hat{\xi}_t)\right) \tag{4.0.58}$$

is continuous, whenever the event $\mathcal{E}_{N_2,\ldots,N_T}$ occurs. Therefore,

$$x_1 \in \mathcal{X}_1 \mapsto \hat{v}_{N_2,\ldots,N_T}(x_1) = F_1(x_1) + \hat{\mathcal{Q}}_2(x_1)$$

is continuous w.p.1, since $F_1(\cdot)$ is continuous. Since $\mathcal{X}_1$ is compact and $\emptyset \neq X_1 \subseteq \mathcal{X}_1$ is closed, it follows that $\mathbb{P}\left[\hat{S}_{N_2,\ldots,N_T} \neq \emptyset\right] = 1$. This completes the proof of the proposition. □

[2]So, we are considering the range $t = 2, \ldots, T - 1$.

Now we derive the exponential rate of convergence for the risk averse multistage setting.

**Proposition 4.0.24.** *Consider a general $T$-stage stochastic programming problem such as (4.0.11), where $T \geq 3$ is an arbitrary integer. Assume that $\phi_t \in \Phi$ satisfies $L(\phi_t) < \infty$, for every $t = 2, \ldots, T$. Assume also that conditions (M0), (M1), and (Mt.1)-(Mt.4) hold, for $t = 1, \ldots, T-1$. Denote the stage sample sizes by $N_2, \ldots, N_T \in \mathbb{N}$, and suppose that the scenario tree is constructed via the* identical conditional sampling *scheme. Then, for $\epsilon > 0$ sufficiently small, the following estimate holds*

$$\mathbb{P}\left[\sup_{x_1 \in X_1} |\hat{v}_{N_2,\ldots,N_T}(x_1) - v(x_1)| \geq \epsilon\right] \leq \sum_{t=1}^{T-1}\Bigg(\exp\{-N_{t+1}\mathsf{m}'_t\} + 2\left[\frac{4\rho D'_t M'_t}{\epsilon/L_t(T-1)}\right]^{n_t+1}\exp\left\{-\frac{N_{t+1}\epsilon^2}{32\kappa^2 L_{t+1}^2\sigma_t^2(T-1)^2}\right\}\Bigg), \tag{4.0.59}$$

*for every $N_2, \ldots, N_T \in \mathbb{N}$, where $L_1 = 1$ and $L_t := L(\phi_t)L_{t-1}$, for $t = 2, \ldots, T$, $M'_t > 0$, $\mathsf{m}'_t \in (0, \infty]$ and $D'_t$ are constants depending on the problem data, for $t = 1, \ldots, T-1$, and $\kappa > 0$ and $\rho > 0$ are absolute constants.*

*Proof.* The proof is similar to the proof of Proposition 2.1.16, although here we apply Proposition 3.3.14 instead of Theorem 2.1.5 for deriving the result. We begin by bounding from above w.p.1 the random quantity

$$\sup_{x_1 \in X_1} |\hat{v}_{N_2,\ldots,N_T}(x_1) - v(x_1)| \tag{4.0.60}$$

by a sum of random variables $\sum_{t=1}^{T-1} L_t Z_t$, where

$$Z_t := \sup_{x_t \in X_t}\left|\hat{\mu}_{\phi_{t+1}}\left(Q_{t+1}(x_t, \hat{\xi}_{t+1})\right) - \mu_{\phi_{t+1}}\left(Q_{t+1}(x_t, \xi_{t+1})\right)\right|, \; t = 1, \ldots, T-1, \tag{4.0.61}$$

$L_1 = 1$, and $L_t = \prod_{s=2}^{t} L(\phi_s)$, for every $t = 2, \ldots, T$. Then, we apply Proposition 3.3.14 for each $Z_t$, $t = 1, \ldots, T-1$, obtaining an upper bound for the probability of $Z_t$ be greater or equal than $\epsilon/L_t(T-1)$ as a function that depends on the problem data and on the sample size $N_{t+1}$.

From (M1) and (Mt.4), $t = 1, \ldots, T-1$, it follows that the event

$$\mathcal{E}_{N_2,\ldots,N_T} := \bigcap_{t=2}^{T}\bigcap_{j=1}^{N_t}\left[\xi_t^j \in E_t\right] \tag{4.0.62}$$

has probability 1, where $E_t$ are the measurable sets appearing in condition (Mt.4), for $t = 1, \ldots, T-1$. We claim that whenever the event $\mathcal{E}_{N_2,\ldots,N_T}$ happens,

$$\sup_{x_1 \in X_1} |\hat{v}_{N_2,\ldots,N_T}(x_1) - v(x_1)| \leq Z_1 + \sum_{t=2}^{T-1} L_t Z_t. \tag{4.0.63}$$

Since $v(x_1) = F_1(x_1) + \mathcal{Q}_2(x_1)$, $\hat{v}_{N_2,\ldots,N_T}(x_1) = F_1(x_1) + \hat{\mathcal{Q}}_2(x_1)$ and $F_1(x_1)$ is finite, for every $x_1 \in X_1$, it follows that

$$
\begin{aligned}
|\hat{v}_{N_2,\ldots,N_T}(x_1) - v(x_1)| &= \left|\hat{\mathcal{Q}}_2(x_1) - \mathcal{Q}_2(x_1)\right|, && (4.0.64)\\
&= \left|\hat{\mu}_{\phi_2}\left(\hat{Q}_2(x_1,\hat{\xi}_2)\right) - \mu_{\phi_2}\left(Q_2(x_1,\xi_2)\right)\right| && (4.0.65)
\end{aligned}
$$

for every $x_1 \in X_1$. Therefore, it is sufficient to bound from above the expression

$$
\sup_{x_1 \in X_1} \left|\hat{\mathcal{Q}}_2(x_1) - \mathcal{Q}_2(x_1)\right|. \tag{4.0.66}
$$

We divide the proof into two steps. In the first one, we show that whenever the event $\mathcal{E}_{N_2,\ldots,N_T}$ occurs the following inequality holds

$$
\sup_{x_t \in \mathcal{X}_t} \left|\hat{\mathcal{Q}}_{t+1}(x_t) - \mathcal{Q}_{t+1}(x_t)\right| \leq Z_t + L(\phi_{t+1}) \sup_{x_{t+1} \in \mathcal{X}_{t+1}} \left|\hat{\mathcal{Q}}_{t+2}(x_{t+1}) - \mathcal{Q}_{t+2}(x_{t+1})\right|, \tag{4.0.67}
$$

for $t = 1, \ldots, T-1$. Let us prove this statement. Take any $x_t \in \mathcal{X}_t$, where $1 \leq t \leq T-1$ is arbitrary. By the triangular inequality, it follows that

$$
\begin{aligned}
\left|\hat{\mathcal{Q}}_{t+1}(x_t) - \mathcal{Q}_{t+1}(x_t)\right| \leq\ & \left|\hat{\mu}_{\phi_{t+1}}\left(Q_{t+1}(x_t,\hat{\xi}_{t+1})\right) - \mu_{\phi_{t+1}}\left(Q_{t+1}(x_t,\xi_{t+1})\right)\right| + \\
& \left|\hat{\mu}_{\phi_{t+1}}\left(\hat{Q}_{t+1}(x_t,\hat{\xi}_{t+1})\right) - \hat{\mu}_{\phi_{t+1}}\left(Q_{t+1}(x_t,\hat{\xi}_{t+1})\right)\right|.
\end{aligned} \tag{4.0.68}
$$

The first term on the right-side of (4.0.68) is less than or equal to $Z_t$. Whenever the event $\mathcal{E}_{N_2,\ldots,N_T}$ happens,

$$
Q_{t+1}(x_t,\xi_{t+1}^j) \in \mathbb{R}, \tag{4.0.69}
$$

for every $j = 1, \ldots, N_{t+1}$. So, $\mu_{\phi_{t+1}}\left(Q_{t+1}(x_t,\hat{\xi}_{t+1})\right) \in \mathbb{R}$ and we can bound the second term applying the inf-sup inequality (see Proposition 2.8.4):

$$
\begin{aligned}
&\left|\hat{\mu}_{\phi_{t+1}}\left(\hat{Q}_{t+1}(x_t,\hat{\xi}_t)\right) - \hat{\mu}_{\phi_{t+1}}\left(Q_{t+1}(x_t,\hat{\xi}_t)\right)\right| = \\
&\qquad \left|\inf_{s\in\mathbb{R}}\left\{s + \tfrac{1}{N_{t+1}}\sum_{j=1}^{N_{t+1}} \phi_{t+1}\left(\hat{Q}_{t+1}(x_t,\xi_{t+1}^j) - s\right)\right\} - \right. \\
&\qquad\qquad \left. \inf_{s\in\mathbb{R}}\left\{s + \tfrac{1}{N_{t+1}}\sum_{j=1}^{N_{t+1}} \phi_{t+1}\left(Q_{t+1}(x_t,\xi_{t+1}^j) - s\right)\right\}\right| \\
&\quad \leq \tfrac{L(\phi_{t+1})}{N_{t+1}} \sup_{s\in\mathbb{R}}\left\{\sum_{j=1}^{N_{t+1}}\left|\left(\hat{Q}_{t+1}(x_t,\xi_{t+1}^j) - s\right) - \left(Q_{t+1}(x_t,\xi_{t+1}^j) - s\right)\right|\right\} \\
&\qquad\qquad = \tfrac{L(\phi_{t+1})}{N_{t+1}}\sum_{j=1}^{N_{t+1}}\left|\hat{Q}_{t+1}(x_t,\xi_{t+1}^j) - Q_{t+1}(x_t,\xi_{t+1}^j)\right|.
\end{aligned} \tag{4.0.70}
$$

As we have also shown in Proposition 2.1.16, we can apply again the inf-sup inequality to obtain

$$\frac{1}{N_{t+1}}\sum_{j=1}^{N_{t+1}}\left|\hat{Q}_{t+1}(x_t,\xi_{t+1}^j)-Q_{t+1}(x_t,\xi_{t+1}^j)\right| \leq \sup_{x_{t+1}\in\mathcal{X}_{t+1}}\left|\hat{\mathcal{Q}}_{t+2}(x_{t+1})-\mathcal{Q}_{t+2}(x_{t+1})\right|, \tag{4.0.71}$$

for every $x_t \in \mathcal{X}_t$. We conclude that (4.0.67) holds for every $t = 1,\ldots,T-1$. Since $\mathcal{Q}_{T+1}(x) = 0 = \hat{\mathcal{Q}}_{T+1}(x)$, for every $x \in \mathbb{R}^{n_T}$, it follows that w.p.1

$$\begin{aligned}\sup_{x_1\in X_1}\left|\mathcal{Q}_2(x_1)-\hat{\mathcal{Q}}_2(x_1)\right| &\leq Z_1 + L(\phi_2)Z_2 + L(\phi_2)L(\phi_3)Z_3 + \cdots + \\ &L(\phi_2)\ldots L(\phi_{T-1})Z_{T-1} = Z_1 + L_2Z_2 + L_3Z_3 + \cdots + L_{T-1}Z_{T-1},\end{aligned} \tag{4.0.72}$$

which proves (4.0.63).

Note that we can apply Proposition 4.0.24 (see also Remark 3.3.17) for every $Z_t$, $t = 1,\ldots,T-1$, since conditions (M1) and (Mt.1)-(Mt.4) are satisfied. Thus, the following bound

$$\mathbb{P}\left[Z_t \geq \frac{\epsilon/L_t}{T-1}\right] \leq \exp\{-N_{t+1}\mathsf{m}_t'\} + 2\left[\frac{4\rho D_t' M_t'}{\epsilon/L_t(T-1)}\right]^{n_t+1}\exp\left\{-\frac{N_{t+1}\epsilon^2}{32\kappa^2L_{t+1}^2\sigma_t^2(T-1)^2}\right\}, \tag{4.0.73}$$

holds, for $\epsilon > 0$ sufficiently small and for all $N_{t+1} \in \mathbb{N}$, $t = 1,\ldots,T-1$. Since

$$\begin{aligned}&\left[\sup_{x_1\in X_1}|\hat{v}_{N_2,\ldots,N_T}(x_1)-v(x_1)|\geq\epsilon\right]\cap\mathcal{E}_{N_2,\ldots,N_T}\subseteq\\ &\left(\bigcup_{t=1}^{T-1}\left[L_tZ_t\geq\frac{\epsilon}{T-1}\right]\right)\cap\mathcal{E}_{N_2,\ldots,N_T}\end{aligned}$$

and $\mathbb{P}\left(\mathcal{E}_{N_2,\ldots,N_T}\right) = 1$, it follows that

$$\begin{aligned}\mathbb{P}\left[\sup_{x_1\in X_1}|\hat{v}_{N_2,\ldots,N_T}(x_1)-v(x_1)|\geq\epsilon\right] &\leq \mathbb{P}\left(\bigcup_{t=1}^{T-1}\left[L_tZ_t\geq\frac{\epsilon}{T-1}\right]\right)\\ &\leq \sum_{t=1}^{T-1}\mathbb{P}\left[L_tZ_t\geq\frac{\epsilon}{T-1}\right]\\ &\leq \sum_{t=1}^{T-1}\Bigg(\exp\{-N_{t+1}\mathsf{m}_t'\} + \\ &2\left[\frac{4\rho D_t'M_t'}{\epsilon/L_t(T-1)}\right]^{n_t+1}\exp\left\{-\frac{N_{t+1}\epsilon^2}{32\kappa^2L_{t+1}^2\sigma_t^2(T-1)^2}\right\}\Bigg).\end{aligned}$$

This completes the proof of the proposition. □

Let us make some remarks about Proposition 4.0.24. Note that it was not necessary to assume conditions (Mt.5), for $t = 1,\ldots,T-1$, in order to derive the exponential bound (4.0.59). Moreover, since $v : X_1 \to \mathbb{R}$ is continuous under the hypotheses of Proposition 4.0.24, we have that $\hat{v}_{N_2,\ldots,N_T}(\cdot)$ is bounded in $X_1$, whenever the event

$$\left[\sup_{x_1 \in X_1} |\hat{v}_{N_2,\ldots,N_T}(x_1) - v(x_1)| < \epsilon\right] \tag{4.0.74}$$

occurs, where $\epsilon > 0$ is arbitrary. So, whenever this event occurs, it automatically follows that $\hat{S}^{\delta}_{N_2,\ldots,N_T} \neq \emptyset$ provided that $\delta > 0$. However, if $\delta = 0$, then it is not clear that $\hat{S}_{N_2,\ldots,N_T} \neq \emptyset$ and some additional regularity conditions such as (Mt.5), $t = 1, \ldots, T-1$, must be assumed in order to guarantee that the SAA problem is solvable.

By changing the definition of the auxiliary sets $\tilde{X}_t$, for $t = 1, \ldots, T-1$, one can obtain an exponential bound like (4.0.59) that works for arbitrarily large values of $\epsilon > 0$ (see Remark 3.3.15). As we have pointed in that remark, this increases the diameter $D'_t$ of the auxiliary sets $\tilde{X}_t$. In the same remark, we point another possibility that does not change the involved constants, but modifies the dependence of the right side of (4.0.59) when $\epsilon$ gets larger than a given threshold $\Delta > 0$ depending on the problem data. For finishing the discussion about this topic, it is worth mentioning that one is usually concerned in how the sample complexity is affected when $\epsilon > 0$ is taken arbitrarily small and not arbitrarily large.

Akin to the static case, given $\epsilon > 0$, $0 \leq \delta < \epsilon$ and $\theta \in (0,1)$, it is possible to obtain sample complexity estimates for a risk averse $T$-stage stochastic programming problem like (4.0.28) applying Proposition 4.0.24. In Corollary 4.0.25 we obtain the sample complexity estimates for the multistage setting.

**Corollary 4.0.25.** *Take any integer $T \geq 3$ and let (4.0.28) be a stochastic programming problem satisfying the assumptions of Proposition 4.0.24. Let $M'_t$, $\mathsf{m}'$, $D'_t$ and $L_t$ be constants depending on the problem data such that (4.0.59) holds for $\epsilon > 0$ sufficiently small and for all $N_2, \ldots, N_T \in \mathbb{N}$. Given $0 < \delta < \epsilon$ and $\theta \in (0,1)$, if $N_t \in \mathbb{N}_{t+1}$ satisfies*

$$N_{t+1} \geq \tfrac{128\kappa^2 L_{t+1}^2 \sigma_t^2 (T-1)^2}{(\epsilon-\delta)^2}\left[(n_{t+1}+1)\log\left(\tfrac{8\rho D'_t M'_t}{(\epsilon-\delta)/L_t(T-1)}\right) + \log\left(\tfrac{4(T-1)}{\theta}\right)\right] \bigvee \left[\tfrac{1}{\mathsf{m}'_t}\log\left(\tfrac{2(T-1)}{\theta}\right)\right], \tag{4.0.75}$$

*for every $t = 1, \ldots, T-1$, then*

$$\mathbb{P}\left(\left[\hat{S}^{\delta}_{N_2,\ldots,N_T} \subseteq S^{\epsilon}\right] \cap \left[\hat{S}^{\delta}_{N_2,\ldots,N_T} \neq \emptyset\right]\right) \geq 1 - \theta. \tag{4.0.76}$$

*If we suppose additionally that conditions (Mt.5) are satisfied, for $t = 1, \ldots, T-1$, then (4.0.76) also holds for $\delta = 0$, whenever $N_{t+1}$ satisfies conditions (4.0.75), for $t = 1, \ldots, T-1$.*

*Proof.* Take any $\epsilon > 0$, $0 \le \delta < \epsilon$ and $\theta \in (0,1)$. It is elementary to verify that whenever $N_{t+1}$ satisfies (4.0.75), it follows that

$$\exp\{-N_{t+1}\mathsf{m}'_t\}+2\left[\frac{8\rho D'_t M'_t}{(\epsilon-\delta)/L_t(T-1)}\right]^{n_t+1}\exp\left\{-\frac{N_{t+1}(\epsilon-\delta)^2}{128\kappa^2 L^2_{t+1}\sigma^2_t(T-1)^2}\right\} \le \frac{\theta}{T-1}, \tag{4.0.77}$$

for $t = 1, \dots, T-1$. By Proposition 4.0.24 we conclude that

$$\mathbb{P}\left[\sup_{x_1 \in X_1} |\hat{v}_{N_2,\dots,N_T}(x_1) - v(x_1)| < \frac{\epsilon-\delta}{2}\right] \ge 1-\theta, \tag{4.0.78}$$

if we take $\epsilon > 0$ sufficiently small. We have already argued (multiple times!) before that

$$\left[\sup_{x_1 \in X_1} |\hat{v}_{N_2,\dots,N_T}(x_1) - v(x_1)| < \frac{\epsilon-\delta}{2}\right] \subseteq \left[\hat{S}^{\delta}_{N_2,\dots,N_T} \subseteq S^{\epsilon}\right]. \tag{4.0.79}$$

Moreover, when $\delta > 0$, the set in the left side of the equation above is also contained in $\left[\hat{S}^{\delta}_{N_2,\dots,N_T} \neq \emptyset\right]$. Assuming conditions (Mt.5), for $t = 1, \dots, T-1$, we also have from Proposition 4.0.23 that $\left[\hat{S}^{\delta}_{N_2,\dots,N_T} \neq \emptyset\right]$ is a set having probability 1, which takes care of the case $\delta = 0$. □

We obtain sample complexity estimates in the multistage risk averse setting that are like the ones obtained in the multistage risk neutral setting. In fact, for multistage risk neutral problems, (4.0.75) becomes

$$\begin{aligned} N_{t+1} \ge \tfrac{128\sigma_t^2(T-1)^2}{(\epsilon-\delta)^2}\left[n_{t+1}\log\left(\tfrac{8\rho D_t M_t}{(\epsilon-\delta)/(T-1)}\right)+\log\left(\tfrac{4(T-1)}{\theta}\right)\right] \bigvee \\ \left[\tfrac{1}{\mathsf{m}_t}\log\left(\tfrac{2(T-1)}{\theta}\right)\right], \end{aligned} \tag{4.0.80}$$

for every $t = 1, \dots, T-1$. Although it is true that $D_t \le D'_t$ and $M_t \le M'_t \le M_t + 1$, the main differences between these two estimates are given by the constants $\kappa$ and $L_t \ge 1$ appearing in the risk averse estimate. Indeed, when one considers the total effect of these constants in the total number of scenarios

$$N = \prod_{t=1}^{T-1} N_t \tag{4.0.81}$$

of the SAA problem, we obtain that

$$\begin{aligned} N_{ra} &\ge N_{rn}\kappa^{T-1}\prod_{t=1}^{T-1} L_{t+1} && (4.0.82)\\ &= N_{rn}\kappa^{T-1}L(\phi_2)^{T-1}L(\phi_3)^{T-2}\dots L(\phi_T). && (4.0.83) \end{aligned}$$

Note that if $L(\phi_t) = L > 1$, for every $t = 2, \ldots, T$, then

$$\prod_{t=2}^{T} L_t = L^{\sum_{t=1}^{T-1} t} = L^{\frac{T(T-1)}{2}}. \tag{4.0.84}$$

This shows that in the risk averse framework the sample complexity of multistage stochastic programming problems can grow much faster with respect to $T$ than in the risk neutral framework.

CHAPTER 5

# A lower bound for the sample complexity of a class of risk neutral dynamic problems

## 5.1 Introduction

Until now we have presented sufficient conditions on the sample sizes $N_2, \ldots, N_T$ that guarantee under some regularity conditions that

$$\mathbb{P}\left(\left[\hat{S}^{\delta}_{N_2,\ldots,N_T} \subseteq S^{\epsilon}\right] \cap \left[\hat{S}^{\delta}_{N_2,\ldots,N_T} \neq \emptyset\right]\right) \geq 1-\theta, \tag{5.1.1}$$

where $\epsilon > 0$, $0 \leq \delta < \epsilon$ and $\theta \in (0,1)$ are the sample complexity parameters. In Section 2.1.2 using the derived sufficient conditions for the sample sizes $N_2, \ldots, N_T$, we noted that in order to obtain a theoretical guarantee like (5.1.1) the total number of scenarios

$$N = \prod_{t=2}^{T} N_t \tag{5.1.2}$$

of the scenario tree used for approximating the true random data explodes exponentially fast with respect to the number of stages $T$. In Section 2.1.2 we have recalled these results for the risk neutral problems. In Chapter 4 we have shown that this behavior gets even worse when one considers the risk averse multistage stochastic programming problems.

One could ask if the sufficient conditions derived for the sample sizes are not too loose in the sense that maybe one could obtain guarantees like (5.1.1) with much smaller sample sizes than the ones prescribed by the sample complexity estimates. Saying equivalently, one could ask if the sample complexity results derived for multistage stochastic programming problems are in some sense tight. One interesting result would be to check if the exponential behavior of the number of scenarios with respect to the number of stages $T$ is really an unavoidable phenomenon for some classes of multistage stochastic programming problems.

In this chapter we answer affirmatively the last question of the previous paragraph. In fact, take any number $0 < r < 2$. We show that there exists an instance of a risk neutral (convex) $T$-stage stochastic programming problem such that in order for (5.1.1) to be satisfied, the total number of scenarios (5.1.2) must be at least

$$\left(\frac{\sigma^2}{\epsilon^{2-r}}\right)^{T-1} (n(T-1))^{T-1} \tag{5.1.3}$$

where $\delta = 0$ and $\theta \in (0, 0.3173)$. It is worth mentioning that the problem instances considered here are "well-behaved" problems in the sense that they satisfy all the regularity conditions considered in Section 2.1.2 and in Chapter 4 to derive the sample complexity estimates for risk neutral (first derived in [69]) and for risk averse problems, respectively. Before we move on, let us mention that the main result of this chapter was published in [53].

For the record here we consider a risk neutral $T$-stage stochastic programming problem

$$\begin{aligned} \min_{x_1 \in X_1} \Bigg\{ f(x_1) := F_1(x_1) + \mathbb{E}\left[\inf_{x_2 \in X_2(x_1,\xi_2)} F_2(x_2,\xi_2)\right. \\ \left.\left. +\mathbb{E}\left[ ... +\mathbb{E}\left[\inf_{x_T \in X_T(x_{T-1},\xi_T)} F_T(x_T,\xi_T)\right]\right]\right]\Bigg\}, \end{aligned} \tag{5.1.4}$$

driven by a stagewise independent random data process $\xi_1, ..., \xi_T$. The others problem components $F_t$ and $X_t$, for $t = 1, \ldots, T$, are as considered before in Section 2.1.2. An instance of (5.1.4) is completely specified by defining the problem components, that also include the specification of the probability distribution of the random data process $\xi_1, ..., \xi_T$. In order to obtain a lower bound for the sample complexity of $T$-stage stochastic problems, we first need to define precisely what we mean by the sample complexity of an instance of a problem. In the sequel we consider the definition of the sample complexity of a class of problems. Afterwards, we show that the results derived in [69] can be seen as an upper bound for the sample complexity of a class of stochastic programming problems. We finish this chapter by deriving a lower bound for the sample complexity of this same class of problems.

Given a problem like (5.1.4) one approximates the random data by constructing a scenario tree through Monte Carlo sampling methods. In order to simplify the exposition, here we consider scenario trees with $T$-levels possessing the following node structure: every $t^{th}$-stage node has $N_{t+1}$ children nodes at level $t+1$, for $t = 1, ..., T-1$. Under this assumption, the total number of scenarios in the tree is equal to

$$N = \prod_{t=2}^{T} N_t.$$

For general scenario trees this does not need to be the case.

As before, we denote the sets of (first-stage) $\epsilon$-optimal solutions, respectively, of the true and the SAA problems as

$$S^{\epsilon} := \{x_1 \in X_1 : f(x_1) \leq v^* + \epsilon\} \tag{5.1.5}$$

and

$$\hat{S}^{\epsilon}_{N_2,\ldots,N_T} := \{x_1 \in X_1 : \hat{f}_{N_2,\ldots,N_T}(x_1) \leq \hat{v}^*_{N_2,\ldots,N_T} + \epsilon\}, \tag{5.1.6}$$

for $\epsilon \geq 0$. The quantities $v^*$ and $\hat{v}^*_{N_2,\ldots,N_T}$ are the optimal values of the true and the SAA problems, respectively.

**Definition 5.1.1.** *(the sample complexity of an instance of $T$-stage stochastic programming problem) Let $(p)$ be an instance of a $T$-stage stochastic programming problem like (5.1.4). Given $\epsilon > 0$, $\delta \in [0, \epsilon)$ and $\theta \in (0, 1)$, we define the set of* viable tuples of samples sizes *as*

$$\mathcal{N}(\epsilon, \delta, \theta; p) := \left\{ (M_2, ..., M_T) : \begin{array}{l} \forall\, (N_2, ..., N_T) \geq (M_2, ..., M_T)\,, \\ \mathbb{P}\left(\left[\hat{S}^{\delta}_{N_2,\ldots,N_T} \subseteq S^{\epsilon}\right] \cap \left[\hat{S}^{\delta}_{N_2,\ldots,N_T} \neq \emptyset\right]\right) \geq 1 - \theta \end{array} \right\}. \tag{5.1.7}$$

*The sample complexity of $(p)$ is defined as*

$$N(\epsilon, \delta, \theta; p) := \inf \left\{ \prod_{t=2}^{T} M_t : (M_2, ..., M_T) \in \mathcal{N}(\epsilon, \delta, \theta; p) \right\}.$$

**Definition 5.1.2.** *(the sample complexity of a class of $T$-stage stochastic programming problems) Let $\mathcal{C}$ be a nonempty class of $T$-stage stochastic programming problems. We define the* sample complexity *of $\mathcal{C}$ as the following quantity depending on the parameters $\epsilon > 0$, $\delta \in [0, \epsilon)$ and $\theta \in (0, 1)$*

$$N(\epsilon, \delta, \theta; \mathcal{C}) := \sup_{p \in \mathcal{C}} N(\epsilon, \delta, \theta; p).$$

**Remark 5.1.3.** *One could have considered the alternative definition*

$$\mathcal{N}^*(\epsilon, \delta, \theta; p) := \left\{ (M_2, ..., M_T) : \mathbb{P}\left(\left[\hat{S}^{\delta}_{M_2,\ldots,M_T} \subseteq S^{\epsilon}\right] \cap \left[\hat{S}^{\delta}_{M_2,\ldots,M_T} \neq \emptyset\right]\right) \geq 1 - \theta \right\} \tag{5.1.8}$$

*instead of ours. In that case, it is clear that $\mathcal{N}^*(\epsilon, \delta, \theta; p) \supseteq \mathcal{N}(\epsilon, \delta, \theta; p)$. Thus,*

$$N^*(\epsilon, \delta, \theta; p) := \inf \left\{ \prod_{t=2}^{T} M_t : (M_2, ..., M_T) \in \mathcal{N}(\epsilon, \delta, \theta; p) \right\} \leq N(\epsilon, \delta, \theta; p), \tag{5.1.9}$$

*for every instance $(p)$ of (5.1.4), $\epsilon > 0$, $\delta \in [0, \epsilon)$ and $\theta \in (0, 1)$. Example 5.1.4 shows that $\mathcal{N}^*(\epsilon, \delta, \theta; p)$ and $\mathcal{N}(\epsilon, \delta, \theta; p)$ could be different. In our definition, if $(N_2, \ldots, N_T) \in N(\epsilon, \delta, \theta; p)$ and $M_t \geq N_t$, for every $t = 2, \ldots, T$, then equation (5.1.1) must also hold for this tuple of sample sizes $(M_2, \ldots, M_T)$. It is worth mentioning that the lower bound estimate that we derive for $N(\epsilon, \delta, \theta; p)$ also holds for $N^*(\epsilon, \delta, \theta; p)$.* □

In order to fix some ideas let us take $T = 2$ for a moment. Although it is well-known that under mild regularity conditions

$$\mathbb{P}\left(\left[\hat{S}^{\delta}_{N_2} \subseteq S^{\epsilon}\right] \cap \left[\hat{S}^{\delta}_{N_2} \neq \emptyset\right]\right) \underset{N_2 \to \infty}{\longrightarrow} 1, \tag{5.1.10}$$

it is not true in general that this sequence of real numbers approaches 1 monotonically (see Example 5.1.4). In particular, it follows that $\mathcal{N}(\epsilon, \delta, \theta; p)$ and $\mathcal{N}^*(\epsilon, \delta, \theta; p)$ does not need to be equal for all values of the sample complexity parameters.

**Example 5.1.4.** *Consider the following* static *stochastic programming problem*

$$\min_{x \in \mathbb{R}} \left\{ f(x) := \mathbb{E}\,|\xi - x| \right\}, \tag{5.1.11}$$

*where $\xi$ is a random variable with finite expected value. It is elementary to show that the set of all medians of $\xi$ is the solution set of this problem. Let us denote the cumulative distribution function of $\xi$ by $H_\xi(z) := \mathbb{P}[\xi \le z]$, for every $z \in \mathbb{R}$. Recall that, by definition, $m \in \mathbb{R}$ is a median of $\xi$ (or of $H_\xi(\cdot)$) if $m$ is a $0.5$-quantile of $\xi$, i.e., if $H_\xi(m) = \mathbb{P}[\xi \le m] \ge 1/2$ and $1 - H_\xi(m-) = \mathbb{P}[\xi \ge m] \ge 1/2$. Moreover, it is well-known that the set of medians of every c.d.f. $H_\xi$ is a nonempty closed bounded interval of $\mathbb{R}$ (see also Proposition 2.3.2).*

*Let $\left\{\xi^1, \ldots, \xi^N\right\}$ be a random sample of $\xi$. The SAA problem is*

$$\min_{x \in \mathbb{R}} \left\{ \hat{f}_N(x) := \hat{\mathbb{E}}\left|\hat{\xi} - x\right| = \frac{1}{N}\sum_{i=1}^{N} \left|\xi^i - x\right| \right\}. \tag{5.1.12}$$

*If $N = 2k - 1$, for some $k \in \mathbb{N}$, then the set of exact optimal solutions for the SAA problem is just $\hat{S}_N = \left\{\xi^{(k)}\right\}$, where $\xi^{(1)} \le \ldots \le \xi^{(N)}$ are the order statistics of the sample $\{\xi^1, \ldots, \xi^N\}$. If $N = 2k$, for some $k \in \mathbb{N}$, then $\hat{S}_N = \left[\xi^{(k)}, \xi^{(k+1)}\right]$. Now it is easy to show that, in general, the sequence of numbers $\{p_N : N \in \mathbb{N}\}$ is not monotone, where*

$$p_N := \mathbb{P}\left(\left[\hat{S}^{\delta}_{N} \subseteq S^{\epsilon}\right] \bigcap \left[\hat{S}^{\delta}_{N} \neq \emptyset\right]\right), \forall N \in \mathbb{N}, \tag{5.1.13}$$

*$\epsilon > 0$ and $\delta \in [0, \epsilon)$ are fixed.*

*We begin by noting that*

$$f(x) = f(0) + \int_0^x \left(2H_\xi(s) - 1\right) ds, \ \forall x \in \mathbb{R}. \tag{5.1.14}$$

*In fact, $f$ is a finite-valued convex function and its right side derivative at $x \in \mathbb{R}$ is equal to $2H(x) - 1$.Thus, (5.1.14) follows from Theorem 2.5.40.*

*Let us assume that $\xi$ is a symmetric random variable around the origin satisfying $\mathbb{P}[\xi \neq 0] > 0$[1]. It follows that $\xi$ and $-\xi$ are equally distributed and that $f(-x) =$*

[1]This is just to rule out the degenerate case $\xi = 0$.

*$\mathbb{E}|\xi + x| = \mathbb{E}|(-\xi) - x| = \mathbb{E}|\xi - x| = f(x)$, for every $x \in \mathbb{R}$. So, $f$ is an even function that assumes its minimum value at the origin. Moreover, by (5.1.14) we see that $f$ is monotonically non-decreasing on $\mathbb{R}_+$, since $H_\xi(s) \geq 1/2$, for all $s \geq 0$. Take $\delta = 0$ and any $\epsilon < 0$. The set of $\epsilon$-solutions for the true problem is $S^\epsilon = [-x^\epsilon, x^\epsilon]$, for some $x^\epsilon > 0$. For $N = 1$, $\left[\hat{S}_N \subseteq S^\epsilon\right]$ if and only if $\left|\xi^1\right| \leq x^\epsilon$. For $N = 2$, $\left[\hat{S}_N \subseteq S^\epsilon\right]$ if and only if $\left|\xi^1\right| \leq x^\epsilon$ and $\left|\xi^2\right| \leq x^\epsilon$. Note also that the SAA problem always has an optimal solution, which implies that $p_N = \mathbb{P}\left(\left[\hat{S}_N^\delta \subseteq S^\epsilon\right]\right)$, for all $N \in \mathbb{N}$. Therefore,*

$$p_2 = \mathbb{P}\left(\left|\xi^1\right| \leq x^\epsilon, \left|\xi^2\right| \leq x^\epsilon\right) = \mathbb{P}\left(\left|\xi^1\right| \leq x^\epsilon\right) \cdot \mathbb{P}\left(\left|\xi^2\right| \leq x^\epsilon\right) = p_1^2 < p_1,$$

*as long as $p_1 < 1$. Of course, this will be the case if we take $\epsilon > 0$ sufficiently small. For a concrete example, just consider $\xi \overset{d}{\sim} U[-1,1]$ and $\epsilon \in (0, 1/2)$. It is elementary to verify that $x^\epsilon = \sqrt{2\epsilon} < 1$, so $p_1 < 1$ and $p_2 < p_1$.* □

In Proposition 5.1.5 we restate Proposition 2.1.17 with minor differences. Here we consider the same regularity conditions (see Page 48) (M0), (M1), (Mt.1)-(Mt.5), for $t = 1, \ldots, T-1$, that were considered in Proposition 2.1.17. Let us recall that in Proposition 2.1.17 we have considered as given real numbers $\tilde{M}_t$, for $t = 1, \ldots, T-1$, satisfying $\tilde{M}_t > M_t = \mathbb{E}\chi_t(\xi_{t+1}) \geq 0$, for every $t = 1, \ldots, T-1$. In order to simplify the exposition, we consider as given an unique real number $\gamma > 1$ and define $\tilde{M}_t := \gamma M_t$, for every $t = 1, \ldots, T-1$. In that case, $\tilde{M}_t > M_t$ if and only if $M_t > 0$. Of course, we can always suppose that $M_t > 0$, for every $t = 1, \ldots, T-1$[2].

**Proposition 5.1.5.** *Consider an instance (p) of a $T$-stage stochastic optimization problem satisfying conditions (M0), (M1) and (Mt.1)-(Mt.5), for $t = 1, \ldots, T-1$. Let $N_2, \ldots, N_T$ be the sample sizes and take any $\gamma > 1$. Suppose that the scenario-tree is constructed via the identical conditional sampling scheme. Then, for $\epsilon > 0$, $\delta \in [0, \epsilon)$ and $\theta \in (0, 1)$, the following inequality is satisfied*

$$N(\epsilon, \delta, \alpha; p) \leq \prod_{t=1}^{T-1} \max\{A_t, B_t\} =: \mathrm{UPPER}(\epsilon, \delta, \theta; p) \tag{5.1.15}$$

[2]If $M_t = 0$, then $\chi_t(\xi_t) = 0$ w.p.1. In that case, condition (Mt.4) would also be satisfied if we take $\chi_t(\xi_t) = M_t > 0$, where $M_t$ is any positive constant. Moreover, it is worth mentioning that if $M_t = 0$, then w.p.1 $Q_{t+1}(x_t, \xi_{t+1}) = Q_{t+1}(x_t', \xi_{t+1})$, for every $x_t, x_t' \in \mathcal{X}_t$, which is a somewhat uninteresting and pathological situation. Indeed, when this is the case, the cost-to-go function does not depend on the random realization $\xi_{t+1}$ and it is indifferent to the choice made in the $t^{th}$-stage.

*where, for* $t = 1, \ldots, T-1$,

$$A_t := \left\lceil \frac{128\sigma_t^2(T-1)^2}{(\epsilon-\delta)^2} \left[ n_t \log\left(\frac{4\rho\gamma D_t M_t (T-1)}{\epsilon-\delta}\right) + \log\left(\frac{4(T-1)}{\theta}\right)\right]\right\rceil, \tag{5.1.16}$$

$$B_t := \left\lceil \frac{1}{I_{\chi_t}(\gamma M_t)} \log\left(\frac{2(T-1)}{\theta}\right)\right\rceil. \tag{5.1.17}$$

*Proof.* Given $\gamma > 1$, define $\tilde{M}_t := \gamma M_t$, for every $t = 1, \ldots, T-1$. Given the sample complexity parameter $\epsilon > 0$, $\delta \in [0, \epsilon)$ and $\theta \in (0,1)$, it follows from Proposition 2.1.17 that the set $\tilde{\mathcal{N}}(\epsilon, \delta, \theta)$ considered in (2.1.143) is contained in $\mathcal{N}(\epsilon, \delta, \theta; p)$. Therefore,

$$\begin{aligned} N(\epsilon, \delta, \theta; p) &\leq \inf\left\{\prod_{t=2}^{T} N_t : (N_2, \ldots, N_T) \in \tilde{\mathcal{N}}(\epsilon, \delta, \theta)\right\} && (5.1.18)\\ &\leq \prod_{t=1}^{T-1} \max\{A_t, B_t\}, && (5.1.19)\end{aligned}$$

where the last inequality follows from Lemma 2.1.18 (see also equations (2.1.159), (2.1.160), (2.1.161) and (2.1.162). □

Take any $\epsilon > 0$, $\delta \in [0, \epsilon)$ and $\theta \in (0,1)$. It is straightforward to derive an upper bound for $N(\epsilon, \delta, \theta; \mathcal{C})$ where $\mathcal{C}$ is the class of all $T$-stage stochastic programming problems like (5.1.4) that satisfy the regularity conditions (M0), (M1), (Mt.1)-(Mt.5), for $t = 1, \ldots, T-1$, and the following uniformly bounded conditions

(UB) There exist positive real constants $\sigma$, $K$, $n \in \mathbb{N}$, $\gamma > 1$ and $\beta$ such that for every instance $(p) \in \mathcal{C}$ and for every $t = 1, \ldots, T-1$, the following assertions hold:

(i) $\sigma_t^2(p) \leq \sigma^2$,

(ii) $D_t(p) \times M_t(p) \leq K$,

(iii) $n_t(p) \leq n$,

(iv) $(0 <)\, \beta \leq I_{\chi_t(p)}(\gamma M_t(p))$.

Then, it is immediate from the previous proposition that

$$N(\epsilon, \delta, \theta; \mathcal{C}) = \sup_{p \in \mathcal{C}} N(\epsilon, \delta, \theta; p) \leq \prod_{t=1}^{T-1} \max\{\bar{A}_t, \bar{B}_t\}, \tag{5.1.20}$$

where

$$\bar{A}_t := \left\lceil \frac{128\sigma^2(T-1)^2}{(\epsilon-\delta)^2}\left[n\log\left(\frac{4\rho\gamma K(T-1)}{\epsilon-\delta}\right) + \log\left(\frac{4(T-1)}{\theta}\right)\right]\right\rceil, \tag{5.1.21}$$

$$\bar{B}_t := \left\lceil \frac{1}{\beta}\log\left(\frac{2(T-1)}{\theta}\right)\right\rceil, \tag{5.1.22}$$

for $t = 1, \ldots, T-1$.

Note that the dependence of $A_t$ and of $\bar{A}_t$ with respect to $\epsilon$ and $\delta$ is given by the difference $\epsilon - \delta > 0$. Therefore, unless stated otherwise, we always take $\delta = 0$ in the sequel. Given a class of problems $\mathcal{C}$ and an instance $(p) \in \mathcal{C}$, we write $N(\epsilon, \theta; \mathcal{C})$ and $N(\epsilon, \theta; p)$, respectively, instead of $N(\epsilon, 0, \theta; \mathcal{C})$ and $N(\epsilon, 0, \theta; p)$. As we have pointed out in Section 2.1.2, for sufficiently small values of $\epsilon > 0$ we have that $A_t \geq B_t$ for each $t = 1, \ldots, T-1$. Therefore, for $\theta \in (0, 1)$ fixed and $\epsilon > 0$ sufficiently small, the order of growth of UPPER$(\cdot)$ with respect to $\epsilon > 0$ is at most

$$\left(\frac{\sigma^2}{\epsilon^2}\left[n\log\left(\frac{K(T-1)}{\epsilon}\right)\right]\right)^{T-1}(T-1)^{2(T-1)}, \tag{5.1.23}$$

where we got rid of the absolute constants in the estimate above. It is worth mentioning that this estimate holds for general multistage stochastic optimization problems and not only for particular subclasses of problems, like the convex or linear subclasses.

## 5.2 The main result

Now, let us derive a *lower bound* for the sample complexity of a class of $T$-stage stochastic problems that satisfies the previous regularity conditions and also condition (UB). Let $T \geq 3$. We consider a family $\mathcal{C} := \{(p_k) : k \in \mathbb{N}\}$ of $T$-stage (convex) stochastic programming problems, where $(p_k)$ is specified by the following data:

(a) $\{\xi_t \overset{d}{\sim} \text{Gaussian}(0,\, s^2 I_n) : t = 2, \ldots, T\}$ is stagewise independent, where $s > 0$ and $n \in \mathbb{N}$,

(b) $F_t^k(x_t, \xi_t) := -2k\,\langle \xi_t, x_t\rangle$, for $t = 2, \ldots, T$,

(c) $X_t^k(x_{t-1}, \xi_t) := \{x_{t-1}\}$, for $t = 2, \ldots, T$,

and $F_1^k(x_1) := \|x_1\|^{2k}$ and $X_1^k := \frac{1}{k}\mathbb{B}_n$, where $\mathbb{B}_n$ is the closed unit Euclidean ball of $\mathbb{R}^n$.

The scenario-tree is constructed following the identical conditional sampling scheme by considering independent random vectors

$$\mathfrak{S}_{N_2,\ldots,N_T} := \left\{ \xi_t^i \stackrel{d}{\sim} \text{Gaussian}(0, s^2 I_n) : \begin{array}{c} t = 2, ..., T, \\ i = 1, ..., N_t \end{array} \right\}. \tag{5.2.1}$$

In this case, the empirical process $\hat{\xi} = \left(\hat{\xi}_1, \ldots, \hat{\xi}_T\right)$ defined on the scenario tree is also stagewise independent.

Let us derive the objective functions $f^k(\cdot)$ and $\hat{f}^k(\cdot)$, respectively, of the true problem and the SAA problem[3] given $\mathfrak{S}_{N_2,\ldots,N_T}$. We begin with the $T^{th}$-stage cost-to-go function obtained by the dynamic programming equation:

$$\begin{aligned} Q_T(x_{T-1}, \xi_T) &= \inf_{x_T \in X_T(x_{T-1}, \xi_T)} \{-2k \langle \xi_T, x_T \rangle\} \\ &= -2k \langle \xi_T, x_{T-1} \rangle . \end{aligned} \tag{5.2.2}$$

The true problem and SAA problem $T^{th}$-stage expected cost-to-go functions are obtained, respectively, by taking the expected value of $\mathcal{Q}_T(x_{T-1}, \xi_T)$ with respect to the true and empirical distribution of $\xi_T$:

$$\begin{aligned} &\mathcal{Q}_T(x_{T-1}) = \mathbb{E}\left[-2k \langle \xi_T, x_{T-1} \rangle\right] = 0, \\ &\hat{\mathcal{Q}}_T(x_{T-1}) = \hat{\mathbb{E}}\left[-2k \left\langle \hat{\xi}_T, x_{T-1} \right\rangle\right] = -2k \left\langle \bar{\xi}_T, x_{T-1} \right\rangle, \end{aligned} \tag{5.2.3}$$

where $\bar{\xi}_T = \frac{1}{N_T} \sum_{i=1}^{N_T} \xi_T^i$. Continuing backward in stages, it is elementary to verify that:

$$\begin{aligned} \mathcal{Q}_t(x_{t-1}, \xi_t) &= -2k \langle \xi_t, x_{t-1} \rangle, \\ \hat{\mathcal{Q}}_t(x_{t-1}, \xi_t) &= -2k \left\langle \bar{\xi}_t + \cdots + \bar{\xi}_T, x_{t-1} \right\rangle, \end{aligned} \tag{5.2.4}$$

where $\bar{\xi}_t := \frac{1}{N_t} \sum_{i=1}^{N_t} \xi_t^i$, for $t = 2, \ldots, T-1$. It follows from (5.2.4) that the true and the SAA first-stage cost-to-go functions are

$$\begin{aligned} &\mathcal{Q}_2(x_1) = 0, \text{ and} \\ &\hat{\mathcal{Q}}_2(x_1) = -2k \left\langle \bar{\xi}_2 + ... + \bar{\xi}_T, x_1 \right\rangle, \end{aligned} \tag{5.2.5}$$

Let us define $\eta := \bar{\xi}_2 + ... + \bar{\xi}_T$. By (5.2.5), it follows that $f^k(x_1) = \|x_1\|^{2k}$ and $\hat{f}^k(x_1) = \|x_1\|^{2k} - 2k \langle \eta, x_1 \rangle$, for $x_1 \in \frac{1}{k}\mathbb{B}_n$. The (unique) first-stage optimal solution of the true problem is $\bar{x}_1 = 0$, so its optimal value is $v^* = 0$. Moreover, the (*exact*) first-stage optimal solution of the SAA problem is given by:

$$\hat{x}_1 = \begin{cases} 0 & \text{, if } \|\eta\| = 0 \\ \dfrac{1}{\|\eta\|^{\gamma_k}} \eta & \text{, if } 0 < \|\eta\| \le (\frac{1}{k})^{2k-1} \\ \dfrac{1}{\|\eta\| \, k} \eta & \text{, if } \|\eta\| > (\frac{1}{k})^{2k-1} \end{cases} \tag{5.2.6}$$

[3]In order to simplify the notation, we have dropped the subscript of the SAA objective function writing $\hat{f}(\cdot)$ instead of $\hat{f}^k_{N_2,\ldots,N_T}(\cdot)$. We proceed in the same way with the cost-to-go functions that we derive in the sequel.

where $\gamma_k = \dfrac{2k-2}{2k-1}$.

Hence, given $\epsilon \in \left(0, \frac{1}{k^{2k}}\right)$, $\hat{x}_1$ is an $\epsilon$-optimal solution of the true problem if and only if $\|\eta\|^{2k(1-\gamma)} \leq \epsilon$. Define $v_k := 2k(1-\gamma_k) = 2k/(2k-1)$. By (5.2.1), $\eta \overset{d}{\sim}$ Gaussian $\left(0, \sum_{t=2}^{T} \dfrac{s^2}{N_t} I_n\right)$. Considering the harmonic mean, say hm, of the numbers $N_2, ..., N_T \in \mathbb{N}$:

$$\frac{T-1}{\text{hm}} := \sum_{t=2}^{T} \frac{1}{N_t},$$

it follows that:

$$\eta \overset{d}{\sim} \text{Gaussian}\left(0, \frac{s^2(T-1)}{\text{hm}} I_n\right). \tag{5.2.7}$$

Let us show that if $(N_2, \ldots, N_T) \in \mathcal{N}(\epsilon, \theta; p_k)$, for $\epsilon \in \left(0, \frac{1}{k^{2k}}\right)$ and $\theta \in (0, \bar{\theta})$, where $\bar{\theta} := \mathbb{P}[\chi_1^2 > 1] \approx 0.3173$, then

$$N := \prod_{t=2}^{T} N_t \geq \left(\frac{s^2}{\epsilon^{2-\frac{1}{k}}}\right)^{T-1} [n(T-1)]^{T-1}. \tag{5.2.8}$$

Since $N(\epsilon, \theta; p_k) = \inf\left\{\prod_{t=2}^{T} N_t : (N_2, \ldots, N_T) \in \mathcal{N}(\epsilon, \theta; p_k)\right\}$, the right side of (5.2.8) will be a lower bound for the sample complexity of the instance $(p_k)$.

Indeed, suppose that $(N_2, \ldots, N_T) \in \mathcal{N}(\epsilon, \theta; p_k)$[4] where $\epsilon$ and $\theta$ are as specified before, then

$$\mathbb{P}[\|\eta\|^{v_k} \leq \epsilon] \geq 1 - \theta. \tag{5.2.9}$$

This is equivalent to $\theta \geq 1 - \mathbb{P}[\|\eta\|^{v_k} \leq \epsilon] = \mathbb{P}[\|\eta\|^{v_k} > \epsilon]$. It follows from (5.2.7) that $\dfrac{\text{hm}}{s^2(T-1)} \|\eta\|^2 \overset{d}{\sim} \chi_n^2$.

Observe also that

$$\mathbb{P}\left[\frac{\text{hm}}{s^2(T-1)} \|\eta\|^2 > \frac{\epsilon^{2/v_k}\,\text{hm}}{s^2(T-1)}\right] = \mathbb{P}[\|\eta\|^{v_k} > \epsilon].$$

Since the sequence $\mathbb{P}[\chi_n^2 > n]$ is monotone increasing and $\mathbb{P}[\chi_1^2 > 1] = \bar{\theta}$, if $\theta \in (0, \bar{\theta})$ we must have that $\frac{\epsilon^{2/v_k}\,\text{hm}}{s^2(T-1)} > n$, i.e.:

$$\text{hm} > \frac{s^2}{\epsilon^{2/v_k}} n(T-1). \tag{5.2.10}$$

It is a well-known result that the harmonic mean of (positive) real numbers is always less than or equal to its *geometric mean*

$$\text{gm} := (N_2 ... N_T)^{1/(T-1)} = N^{1/(T-1)}. \tag{5.2.11}$$

[4]Note that (5.2.9) also holds if we suppose just that $(N_2, \ldots, N_T) \in \mathcal{N}^*(\epsilon, \theta; p_k)$ (see Remark 5.1.3).

So, we arrive at the following lower bound for $N$:

$$\begin{aligned} N &> \left(\frac{s^2 n}{\epsilon^{2-\frac{1}{k}}}\right)^{T-1} (T-1)^{T-1} \\ &= \left(\frac{\sigma_k^2 n}{4\epsilon^{2-\frac{1}{k}}}\right)^{T-1} (T-1)^{T-1}, \end{aligned} \tag{5.2.12}$$

where the last equality follows from the fact that $\sigma_k = 2s$, for all $k \in \mathbb{N}$. Before proving this fact, note the similarities between (5.2.12) and (5.1.23) without the logarithmic term. In particular, observe that the lower bound obtained in (5.2.12) has the factor $(T-1)^{T-1}$ that grows even faster with respect to $T$ than the factorial $T!$. This shows that such multiplicative factor is unavoidable for some problems, and that the number of scenarios for $T$-stage problems can present a much faster order of growth with respect to $T$ than even the exponential one. Moreover, we conclude by (5.2.12) that

$$\lim_{\epsilon \to 0+} \frac{N(\epsilon, \theta, \{p_k : k \in \mathbb{N}\})}{\epsilon^{2(T-1)-s}} = +\infty, \tag{5.2.13}$$

for all $s \in (0, 2(T-1))$, showing a growth order that is almost $1/\epsilon^{2(T-1)}$, when $\epsilon \to 0+$.

Now, let us verify that each instance $(p_k)$ satisfies the regularity conditions (M0), (M1) and (Mt.1)-(Mt.5), for $t = 1, \ldots, T-1$. Conditions (M0) and (M1)[5] are trivially true. Defining $\mathcal{X}_t^k := \frac{1}{k}\mathbb{B}_n$, for $t = 1, \ldots, T-1$, we see that $D_t^k := \operatorname{diam}(\mathcal{X}_t^k) = 2/k$ and $X_t(x_{t-1}, \xi_t) = \{x_{t-1}\} \subseteq \mathcal{X}_t^k$, for every $x_{t-1} \in \mathcal{X}_{t-1}^k$ and $\xi_t \in \mathbb{R}^n$. So, conditions (Mt.1) and (Mt.5) hold, for every $t = 1, \ldots, T-1$. We also have that

$$Q_{t+1}(x_t, \xi_{t+1}) = -2k \{\xi_{t+1}, x_t\}, \tag{5.2.14}$$

for every $x_t \in \mathcal{X}_t^k$ and for every $\xi_{t+1} \in \mathbb{R}^n$. It follows that $\mathcal{Q}_{t+1}(x_t) = \mathbb{E}Q_{t+1}(x_t, \xi_{t+1}) = 0$, for every $x_t \in \mathcal{X}_t$, which shows that conditions (Mt.2) are also satisfied, for $t = 1, \ldots, T-1$. Moreover,

$$|Q_{t+1}(x_t', \xi_{t+1}) - Q_{t+1}(x_t, \xi_{t+1})| = 2k\, |\langle \xi_{t+1}, x_t' - x_t\rangle| \leq 2k\, \|\xi_{t+1}\|\, \|x_t' - x_t\|, \tag{5.2.15}$$

for all $x_t', x_t \in \mathcal{X}_t^k$ and $\xi_{t+1} \in \mathbb{R}^n$, so condition (Mt.3) is satisfied with $\chi_t^k(\xi_{t+1}) = 2k\, \|\xi_{t+1}\|$. Finally, we show that (Mt.3) holds. In fact, for every $x_t \in \mathcal{X}_t^k$

$$Q_{t+1}(x_t, \xi_{t+1}) - \mathcal{Q}_{t+1}(x_t) = -2k \langle \xi_{t+1}, x_t\rangle \overset{d}{\sim} \text{Gaussian}(0, 4k^2 \|x_t\|^2 s^2), \tag{5.2.16}$$

so $\sigma_t = \sigma := 2s > 0$ is such that this family of random variables is $\sigma$-sub-Gaussian.

[5] More precisely, condition (M1) is about the kind of sampling used in order to obtain the SAA problem.

Now, we show that the (UB) condition is also satisfied. It is elementary to verify tha conditions (i.) and (iii.) are satisfied. Moreover,

$$M_t^k := \mathbb{E}\left[\chi_t^k(\xi_{t+1})\right] = 2ks\mathbb{E}\left[\|Z\|\right] = 2kc_n s, \tag{5.2.17}$$

where $Z \overset{d}{\sim}$ Gaussian$(0, I_n)$ and the last equality follows from Lemma 5.2.2 of the Appendix. So, $D_t^k M_t^k = 4c_n s =: K$, for all $k \in \mathbb{N}$ and $t = 1, \ldots, T-1$. Since $\|\xi\| \leq \|\xi\|_1$, we obtain the following estimate[6]

$$M_{\chi_k}(r) = \exp\left(2k\|\xi\|\theta\right) \leq 2^n \exp\left(2nk^2s^2r^2\right), \forall r \in \mathbb{R}. \tag{5.2.18}$$

Consequently, for $\gamma > 1$ (see also Lemma 5.2.3)

$$I_{\chi_t^k}(\gamma M_k) \geq \frac{1}{4}\gamma^2 - n\log(2), \tag{5.2.19}$$

for all $t = 1, \ldots, T-1$ and $k \in \mathbb{N}$. Taking $\gamma = 2\sqrt{n}$ we obtain that $I_{\chi_t^k}(\gamma L_k) \geq n(1 - \log(2)) \geq 1 - \log(2) =: \beta (> 0)$. So, we have shown that all items of (UB) are satisfied. We can summarize the discussion above in the following proposition.

**Proposition 5.2.1.** *Let $\mathcal{C}$ be the class of all $T$-stage stochastic convex problems satisfying the regularity conditions (M0), (M1), (Mt.1)-(Mt.5), for $t = 1, \ldots, T-1$, and (UB) with arbitrary constants $\sigma > 0$, $M > 0$, $n \in \mathbb{N}$, $\gamma > 1$ and $\beta > 0$, where $\frac{1}{2}\gamma^2 \geq \beta + n\log(2)$. Then, for $\theta \in (0,1)$ sufficiently small,*

$$\lim_{\epsilon \to 0+} \frac{N(\epsilon, \theta; \mathcal{C})}{\epsilon^{2(T-1)-r}} = +\infty, \text{ for all } r \in (0, 2(T-1)). \tag{5.2.20}$$

The proof is immediate, since $\mathcal{C} \supseteq \{p_k : k \in \mathbb{N}\}$, for sufficiently small $s > 0$, which implies that $N(\epsilon, \theta; \mathcal{C}) \geq N(\epsilon, \theta; \{p_k : k \in \mathbb{N}\})$.

# Some lemmas

**Lemma 5.2.2.** *Let $\xi$ be a multivariate standard Gaussian random vector, i.e. $\xi \overset{d}{\sim}$ Gaussian$(0, I_n)$, $n \in \mathbb{N}$. Then*

$$\frac{n}{\sqrt{n+1}} \leq \mathbb{E}\|\xi\| = \frac{\sqrt{2}\,\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \leq \sqrt{n}, \tag{5.2.21}$$

*where $\Gamma(s) := \int_0^{+\infty} u^{s-1}\exp\{-u\}du$, $s > 0$, is the gamma function.*

[6]In particular, this estimate shows that $M_{\chi_k}(r) < \infty$, for every $r \in \mathbb{R}$.

*Proof.* The probability density function of $\xi$ is equal to

$$h_\xi(x) = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{\|x\|^2}{2}\right\}, \forall x \in \mathbb{R}^n. \tag{5.2.22}$$

Observe that $h_\xi(x) = g(\|x\|)$, where

$$g(r) = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{r^2}{2}\right\}, \forall r \geq 0.$$

So, the expected value of $\|\xi\|$ is equal to

$$\mathbb{E}\,\|\xi\| = \int_{x\in\mathbb{R}^n} \|x\|\, g(\|x\|)dx = \int_0^{+\infty} S_{n-1}(r) r g(r) dr, \tag{5.2.23}$$

where $S_{n-1}(r) = \dfrac{n\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)} r^{n-1}$ is the surface area of the sphere of $\mathbb{R}^n$ with radius $r$. So, we need to solve the following integral in one variable

$$\mathbb{E}\,\|\xi\| = \frac{n\pi^{n/2}}{(2\pi)^{n/2}\Gamma\left(\frac{n}{2}+1\right)} \int_0^{+\infty} r^n \exp\left\{-r^2/2\right\} dr. \tag{5.2.24}$$

Making the change of variables $u = r^2/2$, it is elementary to verify the equality in (5.2.21). The upper bound is an immediate consequence of Jensen's inequality, since $\mathbb{E}\,\|\xi\|^2 = n$. Finally, using an induction argument on $k \in \mathbb{N}$, for $n = 2k-1$ and for $n = 2k$ (separately), one can show the lower bound after some tedious calculations. It is not difficult to verify our claims and, for such, it is worth noting that $\Gamma(s) = (s-1)\Gamma(s-1)$, for $s > 1$, and $\Gamma(1/2) = \sqrt{\pi}$. □

**Lemma 5.2.3.** *Let* $\chi_k(\xi) := 2k\,\|\xi\|$, *where* $\xi \overset{d}{\sim} \text{Gaussian}(0, s^2 I_n)$, $s > 0$, $c_n = \sqrt{2}\,\Gamma\left(\frac{n+1}{2}\right)/\Gamma\left(\frac{n}{2}\right)$ *and* $k \in \mathbb{N}$. *The following conditions hold:*

*i.* $M_k := \mathbb{E}\chi_k(\xi) = 2kc_n s$, $\forall k \in \mathbb{N}$.

*ii.* $I_{\chi_k}(\gamma M_k) \geq \frac{1}{4}\gamma^2 - n\log(2)$, $\forall k \in \mathbb{N}$ *and* $\gamma > 1$.

*Proof.* Item (i) follows immediately from Lemma 5.2.2. Let us item (ii). Taking the logarithm in (5.2.18), we obtain

$$m_{\chi_k}(r) := \log\left(M_{\chi_k}(r)\right) \leq n\log(2) + 2nk^2 s^2 r^2, \forall r \in \mathbb{R}. \tag{5.2.25}$$

Let $y > 0$ be arbitrary. Then

$$\begin{aligned} I_{\chi_k}(y) &= \sup_{r\in\mathbb{R}} \{ry - m_{\chi_k}(r)\} \\ &\geq \sup_{r\in\mathbb{R}} \{ry - n\log(2) - 2nk^2 s^2 r^2\} \\ &= \frac{y^2}{8ns^2k^2} - n\log(2). \end{aligned} \tag{5.2.26}$$

Given $\gamma > 1$, take $y := \gamma M_k$ on (5.2.26) in order to obtain the following lower bound

$$I_{\chi_k}(\gamma M_k) \geq \frac{c_n^2 \gamma^2}{2n} - n\log(2). \tag{5.2.27}$$

From (5.2.21), it follows that $\frac{c_n^2}{n} \geq \frac{n}{n+1} \geq 1/2$, for all $n \in \mathbb{N}$. This completes the proof of the lemma.. $\square$

# Bibliography


[1] C. Acerbi, D. Tasche, On the coherence of expected shortfall, Journal of Banking and Finance, 26 (2002) 1491–1507.

[2] A. Araujo, E. Guiné, The central limit theorem for real and Banach valued random variables. Wiley, 1980.

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

[4] T.G. Bailey, P. Jensen, D. Morton, Response surface analysis of two-stage stochastic linear programming with recourse, Naval Research Logistics 46 (1999) 753–778.

[5] A. Ben-Tal, M. Teboulle, An old-new concept of convex risk measures: the optimized certainty equivalent, Mathematical Finance 17 (2007) 449–476.

[6] C. Berge, Topological spaces: including a treatment of multi-valued functions, vector spaces and convexity. Oliver and Boyd, 1963.

[7] D.P. Bertsekas, Convex optimization theory. Athena Scientific, 2009.

[8] P. Billingsley, Probability and measure. Wiley, 1995.

[9] J.R. Birge, F. Louveaux, Introduction to stochastic programming. Springer, 2011.

[10] J.F. Bonnans, A. Shapiro, Perturbation analysis of optimization problems. Springer, 2000.

[11] V. Brazauskas, B.L. Jones, M.L. Puri, R. Zitikis, Estimating conditional tail expectation with actuarial applications in view, Journal of Statistical Planning and Inference 138 (2008) 3590–3640.

[12] D.B. Brown, Large deviations bounds for estimating conditional value-at-risk, Operations Research Letters 35 (2007) 722–730.

[13] R.S. Burachik, A.N. Iusem, Set-valued mappings and enlargements of monotone operators. Springer, 2008.

[14] K.L. Chung, A course in probability theory. Academic Press, 2001.

[15] G. Dantzig, G. Infanger, Multi-stage stochastic linear programs for portfolio optimization, Annals of Operations Research 45 (1993) 59–76.

[16] V.L. de Matos, D.P. Morton, E.C. Finardi, Assessing policy quality in a multistage stochastic program for long-term hydrothermal scheduling, Annals of Operations Research (2016) 1–19.

[17] P. Delbaen, Coherent risk measures on general probability spaces, in: P.J. Schönbucher, K. Sandmann (Eds.), Advances in Finance and Stochastics: Essays in Honour of Dieter Sondermann, Springer, 2002.

[18] A. Dembo, O. Zeiouni, Large deviations techniques and applications. Springer, 1998.

[19] J. Dupačová, V. Kozmík, Structure of risk-averse multistage stochastic stochastic programs, OR Spectrum 37 (2015) 559–582.

[20] J. Dupačová, J. Polívka, Asset-liability management for Czech pension funds using stochastic programming, Annals of Operations Research 165 (2009) 5–28.

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

[22] R. Durrett, Probability: theory and examples. Cambridge University Press, 2010.

[23] M. Evans, T. Swartz, Approximating integrals via Monte Carlo and deterministic methods. Oxford, 2000.

[24] B. Fhoula, A. Hajji, M. Rekik, Stochastic dual dynamic programming or transportation planning under demand uncertainty, in 2013 International Conference on Advanced Logistics and Transport (May 2013) 550–555.

[25] H. Föllmer, A. Schied, Convex measures of risk and trading constraints, Finance and Stochastics 6 (2002) 429–447.

[26] V. Guigues, C. Sagastizábal, The value of rolling horizon policies for risk-averse hydro-thermal planning, European Journal of Operational Research 217 (2012) 219–240.

[27] Y.T. Herer, M. Tzur, E. Yücesan, The multilocation transshipment problem, IIE Transaction 38 (2006) 185–200.

[28] J. Higle, S. Sen, Stochastic decomposition: a statistical method for large scale stochastic linear programming. Kluwer, 1996.

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

[30] T. Homem-de-Mello, G. Bayraksan, Monte Carlo sampling-based methods for stochastic optimization, Surveys in Operations Research and Management Science 19 (2014) 56–85.

[31] T. Homem-de-Mello and B. Pagnoncelli, Risk aversion in multistage stochastic programming : A modeling and algorithmic perspective, European Journal of Operational Research, 249 (2016) 188–199.

[32] B.L. Jones, R. Zitikis, Empirical estimation of risk measures and related quantities, North American Actuarial Journal 7 (2003) 44–54.

[33] B.L. Jones, R. Zitikis, Risk measures, distortion parameters, and their empirical estimation, Insurance: Mathematics and Economics 41 (2007) 279–297.

[34] J.-P. Kahane, Propriétés locales des fonctions à séries de Fourier aléatoires. Studia Mathematica 19 (1960) 1–25.

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

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

[37] A.J. Kleywegt, A. Shapiro, T. Homem-De-Mello, The sample average approximation method for stochastic discrete optimization, SIAM Journal on Optimization 12 (2001) 479–502.

[38] V. Kozmík, D. Morton, Risk-averse stochastic dual dynamic programming, Technical report, University of Texas at Austin, 2013. Available at: `http://www.optimization-online.org`, 2013.

[39] V. Kozmík, D. Morton, Evaluating policies in risk-averse multi-stage stochastic programming, Mathematical Programming 152 (2015) 275–300.

[40] J.T. Linderoth, A. Shapiro, S.J. Wright, The empirical behavior of sampling methods for stochastic programming, Annals of Operations Research 142 (2006) 215–241.

[41] H.M. Markowitz, Portfolio selection, Journal of Finance 7 (1952) 77–91.

[42] H.M. Markowitz, Portfolio selection. Wiley, 1959.

[43] H.M. Markowitz, Mean-variance analysis in portfolio choice and capital markets, Blackwell, 1987.

[44] C. McDiarmid, On the method of bounded differences, in: J. Siemons (Ed.), Surveys in Combinatorics, Cambridge University Press (1989) 148–188.

[45] T. Pennanen, Epi-convergent discretizations of multistage stochastic programs, Mathematics of Operations Research 30 (2005) 245–256.

[46] M.V.F. Pereira, L.M.V.G. Pinto, Multi-stage stochastic optimization applied to energy planning, Mathematical Programming 52 (1991) 359–375.

[47] G.C. Pflug, Some remarks on the value-at-risk and the conditional value-at-risk, in: S. Uryasev (Ed.), Probabilistic constrained optimization: methodology and applications, Kluwer (2000) 272–281.

[48] G.C. Pflug, A. Pichler, Multistage stochastic optimization. Springer, 2014.

[49] G.C. Pflug, W. Römisch, Modeling, measuring and managing risk. World Scientific, 2007.

[50] G.C. Pflug, N. Wozabal, Asymptotic distribution of law-invariant risk functionals, Finance and Stochastics 14 (2010) 397–418.

[51] A. Philpott, V.L. de Matos, Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion, European Journal of Operational Research, 218 (2012) 470–483.

[52] E.L. Plambeck, B.R. Fu, S.M. Robinson, R. Suri, Sample-path optimization of convex stochastic performance functions, Mathematical Programming, Series B, 75 (1996) 137–176.

[53] M.M.C.R. Reaiche, A note on sample complexity of multistage stochastic programs, Operations Research Letters 44 (2016) 430–435.

[54] O. Rivasplata, Subgaussian random variables: an expository note, E-print available at: `http://www.math.ualberta.ca/~orivasplata/publications/subgaussians.pdf`, 2012.

[55] H. Robbins, S. Monro, On a stochastic approximation method, Annals of Mathematical Statistics 22 (1951) 400–407.

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

[57] R.T. Rockafellar, S. Uryasev, Optimization of conditional value at risk, The Journal of Risk 2 (2000) 21–41.

[58] R.T. Rockafellar, S. Uryasev, M. Zabarankin, Generalized deviations in risk analysis, Finance and Stochastics 10 (2006) 51–74.

[59] R.T. Rockafellar, R.J.-B. Wets, Variational analysis. Springer, 2009.

[60] H.L. Royden, Real analysis. Macmillan, 1988.

[61] R.Y. Rubinstein, A. Shapiro, Discrete event systems: sensitivity analysis and stochastic optimization by the score function method. John Wiley & Sons, 1993.

[62] A. Ruszczyński, Risk-averse dynamic programming for Markov decision processes, Mathematical Programming (Series B), 125 (2010) 235–261.

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

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

[65] A. Shapiro, Asymptotic behavior of optimal solutions in stochastic programming, Mathematics of Operations Research 18 (1993) 829–845.

[66] A. Shapiro, Statistical inference of stochastic optimization problems, in: S. Uryasev (Ed.), Probabilistic constrained optimization: methodology and applications, Kluwer (2000) 282–304.

[67] A. Shapiro, Inference of statistical bounds for multistage stochastic programming problems, Mathematical Methods of Operations Research 58 (2003) 57–68.

[68] A. Shapiro, Monte Carlo sampling methods, in: A. Ruszcyński and A. Shapiro (Eds.), Stochastic Programming, Elsevier (2003) 353–425.

[69] A. Shapiro, On complexity of multistage stochastic programs, Operations Research Letters 34 (2006) 1–8.

[70] A. Shapiro, Analysis of stochastic dual dynamic programming method, European Journal of Operational Research 209 (2011) 63–72.

[71] A. Shapiro, Consistency of sample estimates of risk averse stochastic programs, Journal of Applied Probability 50 (2013) 533–541.

[72] A. Shapiro, F.G. Cabral, J.P. da Costa, Guidelines for choosing parameters $\lambda$ and $\alpha$ for risk averse approach, Technical Report of the Cooperation between Georgia Institute of Technology and Operador Nacional do Sistema, 2015.

[73] A. Shapiro, D. Dentcheva, A. Ruszczyński, Lectures on stochastic programming: modeling and theory. SIAM, 2014.

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

[75] A. Shapiro, A. Nemirovski, On complexity of stochastic programming problems, in: V. Jeyakumar and A.M. Rubinov (Eds.), Continuous Optimization: Current Trends and Applications, Springer (2005) 111–144.

[76] A. Shapiro, W. Tekaya, J.P. da Costa, M.P. Soares, Risk neutral and risk averse stochastic dual dynamic programming method, European Journal of Operational Research 224 (2013) 375–391.

[77] I.H. Sloan, S. Joe, Lattice methods for multiple integration. Oxford University, 1994.

[78] W. van Ackooij, W. de Oliveira, Y. Song, On regularization with normal solutions in decomposition methods for multistage stochastic programming, E-print available at: `http://www.optimization-online.org`, 2017.

[79] R. Vershynin, Introduction to the non-asymptotic analysis of random matrices, in: Y. Yeldar and G. Kutynik (Eds.), Compressed sensing: theory and applications, Cambridge University Press (2012) 210–268.

[80] D. Wozabal, N. Wozabal, Asymptotic consistency of risk functionals, Journal of Nonparametrics Statistics 21 (2009) 977–990.