Chapter 6

Risk Averse Optimization

Andrzej Ruszczyński and Alexander Shapiro

6.1 Introduction

So far, we have discussed stochastic optimization problems, in which the objective function was defined as the expected value $f(x) := \mathbb{E}[F(x, \omega)]$. The function $F : \mathbb{R}^n \times \Omega \to \mathbb{R}$ models the random outcome, for example, the random cost, and is assumed to be sufficiently regular so that the expected value function is well defined. For a feasible set $X \subset \mathbb{R}^n$, the stochastic optimization model

$$\operatorname*{Min}_{x \in X} f(x) \tag{6.1}$$

optimizes the random outcome $F(x, \omega)$ *on average*. This is justified when the Law of Large Numbers can be invoked and we are interested in the long-term performance, irrespective of the fluctuations of specific outcome realizations. The shortcomings of such an approach can be clearly illustrated by the example of portfolio selection discussed in section 1.4. Consider problem (1.34) of maximizing the expected return rate. Its optimal solution suggests concentrating on investment in the assets having the highest expected return rate. This is not what we would consider reasonable, because it leaves out all considerations of the involved risk of losing all invested money. In this section we discuss stochastic optimization from a point of view of risk averse optimization.

A classical approach to risk averse preferences is based on the expected utility theory, which has its roots in mathematical economics (we touched on this subject in section 1.4). In this theory, in order to compare two random outcomes we consider expected values of some scalar transformations $u : \mathbb{R} \to \mathbb{R}$ of the realization of these outcomes. In a minimization problem, a random outcome Z_1 (understood as a scalar random variable) is preferred over a random outcome Z_2 if

$$\mathbb{E}[u(Z_1)] < \mathbb{E}[u(Z_2)].$$

The function $u(\cdot)$, called the *disutility function*, is assumed to be nondecreasing and convex. Following this principle, instead of problem (6.1), we construct the problem

$$\operatorname*{Min}_{x \in X} \mathbb{E}[u(F(x, \omega))]. \tag{6.2}$$

Observe that it is still an expected value problem, but the function F is replaced by the composition $u \circ F$. Since $u(\cdot)$ is convex, we have by Jensen's inequality that

$$u(\mathbb{E}[F(x, \omega)]) \le \mathbb{E}[u(F(x, \omega))].$$

That is, a sure outcome of $\mathbb{E}[F(x, \omega)]$ is at least as good as the random outcome $F(x, \omega)$. In a maximization problem, we assume that $u(\cdot)$ is concave (and still nondecreasing). We call it a *utility function* in this case. Again, Jensen's inequality yields the preference in terms of expected utility:

$$u(\mathbb{E}[F(x, \omega)]) \ge \mathbb{E}[u(F(x, \omega))].$$

One of the basic difficulties in using the expected utility approach is specifying the utility or disutility function. They are very difficult to elicit; even the authors of this book cannot specify their utility functions in simple stochastic optimization problems. Moreover, using some arbitrarily selected utility functions may lead to solutions which are difficult to interpret and explain. A modern approach to modeling risk aversion in optimization problems uses the concept of risk measures. These are, generally speaking, functionals which take as their argument the entire collection of realizations $Z(\omega) = F(x, \omega)$, $\omega \in \Omega$, understood as an object in an appropriate vector space. In the following sections we introduce this concept.

6.2 Mean–Risk Models

6.2.1 Main Ideas of Mean–Risk Analysis

The main idea of mean–risk models is to characterize the uncertain outcome $Z_x(\omega) = F(x, \omega)$ by two scalar characteristics: the *mean* $\mathbb{E}[Z]$, describing the expected outcome, and the *risk (dispersion measure)* $\mathbb{D}[Z]$, which measures the uncertainty of the outcome. In the mean–risk approach, we select from the set of all possible solutions those that are *efficient*: for a given value of the mean they minimize the risk, and for a given value of risk they maximize the mean. Such an approach has many advantages: it allows one to formulate the problem as a parametric optimization problem and it facilitates the trade-off analysis between mean and risk.

Let us describe the mean–risk analysis on the example of the minimization problem (6.1). Suppose that the risk functional is defined as the variance $\mathbb{D}[Z] := \mathbb{V}\mathrm{ar}[Z]$, which is well defined for $Z \in \mathcal{L}_2(\Omega, \mathcal{F}, P)$. The variance, although not the best choice, is easiest to start from. It is also important in finance. Later in this chapter we discuss in much detail desirable properties of the risk functionals.

In the mean–risk approach, we aim at finding efficient solutions of the problem with two objectives, namely, $\mathbb{E}[Z_x]$ and $\mathbb{D}[Z_x]$, subject to the feasibility constraint $x \in X$. This can be accomplished by techniques of multiobjective optimization. Most convenient, from our perspective, is the idea of *scalarization*. For a coefficient $c \geq 0$, we form a composite objective functional

$$\rho[Z] := \mathbb{E}[Z] + c\mathbb{D}[Z]. \tag{6.3}$$

The coefficient c plays the role of the price of risk. We formulate the problem

$$\operatorname*{Min}_{x \in X} \; \mathbb{E}[Z_x] + c\mathbb{D}[Z_x]. \tag{6.4}$$

By varying the value of the coefficient c, we can generate in this way a large ensemble of efficient solutions. We already discussed this approach for the portfolio selection problem, with $\mathbb{D}[Z] := \mathbb{V}\mathrm{ar}[Z]$, in section 1.4.

An obvious deficiency of variance as a measure of risk is that it treats the excess over the mean equally as the shortfall. After all, in the minimization case, we are not concerned if a particular realization of Z is significantly below its mean; we do not want it to be too large. Two particular classes of risk functionals, which we discuss next, play an important role in the theory of mean–risk models.

6.2.2 Semideviations

An important group of risk functionals (representing dispersion measures) are *central semideviations*. The *upper* semideviation of order p is defined as

$$\sigma_p^+[Z] := \Big(\mathbb{E}\Big[\big(Z - \mathbb{E}[Z]\big)_+^p\Big]\Big)^{1/p}, \tag{6.5}$$

where $p \in [1, \infty)$ is a fixed parameter. It is natural to assume here that considered random variables (uncertain outcomes) $Z : \Omega \to \mathbb{R}$ belong to the space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$, i.e., that they have finite pth order moments. That is, $\sigma_p^+[Z]$ is well defined and *finite* valued for all $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$. The corresponding mean–risk model has the general form

$$\operatorname*{Min}_{x \in X} \; \mathbb{E}[Z_x] + c\sigma_p^+[Z_x]. \tag{6.6}$$

The upper semideviation measure is appropriate for minimization problems, where $Z_x(\omega) = F(x, \omega)$ represents a cost, as a function of $\omega \in \Omega$. It is aimed at penalization of an excess of Z_x over its mean. If we are dealing with a maximization problem, where Z_x represents some reward or profit, the corresponding risk functional is the *lower* semideviation

$$\sigma_p^-[Z] := \Big(\mathbb{E}\Big[\big(\mathbb{E}[Z] - Z\big)_+^p\Big]\Big)^{1/p}, \tag{6.7}$$

where $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$. The resulting mean–risk model has the form

$$\operatorname*{Max}_{x \in X} \; \mathbb{E}[Z_x] - c\sigma_p^-[Z_x]. \tag{6.8}$$

In the special case of $p = 1$, both left and right first order semideviations are related to the mean absolute deviation

$$\sigma_1(Z) := \mathbb{E}\big|Z - \mathbb{E}[Z]\big|. \tag{6.9}$$

Proposition 6.1. *The following identity holds:*

$$\sigma_1^+[Z] = \sigma_1^-[Z] = \tfrac{1}{2}\sigma_1[Z], \quad \forall Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P). \tag{6.10}$$

Proof. Denote by $H(\cdot)$ the cumulative distribution function (cdf) of Z and let $\mu := \mathbb{E}[Z]$. We have

$$\sigma_1^-[Z] = \int_{-\infty}^{\mu} (\mu - z)\, dH(z) = \int_{-\infty}^{\infty} (\mu - z)\, dH(z) - \int_{\mu}^{\infty} (\mu - z)\, dH(z).$$

The first integral on the right-hand side is equal to 0, and thus $\sigma_1^-[Z] = \sigma_1^+[Z]$. The identity (6.10) follows now from the equation $\sigma_1[Z] = \sigma_1^-[Z] + \sigma_1^+[Z]$. $\square$

We conclude that using the mean absolute deviation instead of the semideviation in mean–risk models has the same effect, just the parameter c has to be halved. The identity (6.10) does not extend to semideviations of higher orders, unless the distribution of Z is symmetric.

6.2.3 Weighted Mean Deviations from Quantiles

Let $H_Z(z) = \Pr(Z \le z)$ be the cdf of the random variable Z and $\alpha \in (0, 1)$. Recall that the *left-side α-quantile* of H_Z is defined as

$$H_Z^{-1}(\alpha) := \inf\{t : H_Z(t) \ge \alpha\} \tag{6.11}$$

and the *right-side α-quantile* as

$$\sup\{t : H_Z(t) \le \alpha\}. \tag{6.12}$$

If Z represents losses, the (left-side) quantile $H_Z^{-1}(1-\alpha)$ is also called *Value-at-Risk* and denoted $\mathsf{V@R}_\alpha(Z)$, i.e.,

$$\mathsf{V@R}_\alpha(Z) = H_Z^{-1}(1-\alpha) = \inf\{t : \Pr(Z \le t) \ge 1-\alpha\} = \inf\{t : \Pr(Z > t) \le \alpha\}.$$

Its meaning is the following: *losses larger than* $\mathsf{V@R}_\alpha(Z)$ *occur with probability not exceeding* α. Note that

$$\mathsf{V@R}_\alpha(Z+\tau) = \mathsf{V@R}_\alpha(Z) + \tau, \quad \forall \tau \in \mathbb{R}. \tag{6.13}$$

The weighted mean deviation from a quantile is defined as follows:

$$\mathsf{q}_\alpha[Z] := \mathbb{E}\left[\max\left\{(1-\alpha)(H_Z^{-1}(\alpha) - Z), \alpha(Z - H_Z^{-1}(\alpha))\right\}\right]. \tag{6.14}$$

The functional $\mathsf{q}_\alpha[Z]$ is well defined and finite valued for all $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$. It can be easily shown that

$$\mathsf{q}_\alpha[Z] = \min_{t\in\mathbb{R}} \left\{\varphi(t) := \mathbb{E}\big[\max\{(1-\alpha)(t - Z), \alpha(Z - t)\}\big]\right\}. \tag{6.15}$$

Indeed, the right- and left-side derivatives of the function $\varphi(\cdot)$ are

$$\varphi'_+(t) = (1-\alpha)\Pr[Z \le t] - \alpha\Pr[Z > t],$$
$$\varphi'_-(t) = (1-\alpha)\Pr[Z < t] - \alpha\Pr[Z \ge t].$$

At the optimal t the right derivative is nonnegative and the left derivative nonpositive, and thus

$$\Pr[Z < t] \le \alpha \le \Pr[Z \le t].$$

This means that every α-quantile is a minimizer in (6.15).

The risk functional $\mathsf{q}_\alpha[\,\cdot\,]$ can be used in mean–risk models, both in the case of minimization

$$\operatorname*{Min}_{x\in X}\ \mathbb{E}[Z_x]+c\mathsf{q}_{1-\alpha}[Z_x] \tag{6.16}$$

and in the case of maximization

$$\operatorname*{Max}_{x\in X}\ \mathbb{E}[Z_x]-c\mathsf{q}_{\alpha}[Z_x]. \tag{6.17}$$

We use $1-\alpha$ in the minimization problem and α in the maximization problem, because in practical applications we are interested in these quantities for small α.

6.2.4 Average Value-at-Risk

The mean-deviation from quantile model is closely related to the concept of Average Value-at-Risk.[39] Suppose that Z represents losses and we want to satisfy the chance constraint:

$$\mathsf{V@R}_\alpha[Z_x]\le 0. \tag{6.18}$$

Recall that

$$\mathsf{V@R}_\alpha[Z]=\inf\{t:\Pr(Z\le t)\ge 1-\alpha\},$$

and hence constraint (6.18) is equivalent to the constraint $\Pr(Z_x\le 0)\ge 1-\alpha$. We have that[40] $\Pr(Z_x>0)=\mathbb{E}\left[\mathbf{1}_{(0,\infty)}(Z_x)\right]$, and hence constraint (6.18) can also be written as the expected value constraint:

$$\mathbb{E}\left[\mathbf{1}_{(0,\infty)}(Z_x)\right]\le\alpha. \tag{6.19}$$

The source of difficulties with probabilistic (chance) constraints is that the step function $\mathbf{1}_{(0,\infty)}(\cdot)$ is not convex and, even worse, it is discontinuous at zero. As a result, chance constraints are often nonconvex, even if the function $x\mapsto Z_x$ is convex almost surely. One possibility is to approach such problems by constructing a convex approximation of the expected value on the left of (6.19).

Let $\psi:\mathbb{R}\to\mathbb{R}$ be a nonnegative valued, nondecreasing, convex function such that $\psi(z)\ge\mathbf{1}_{(0,\infty)}(z)$ for all $z\in\mathbb{R}$. By noting that $\mathbf{1}_{(0,\infty)}(tz)=\mathbf{1}_{(0,\infty)}(z)$ for any $t>0$ and $z\in\mathbb{R}$, we have that $\psi(tz)\ge\mathbf{1}_{(0,\infty)}(z)$ and hence the following inequality holds:

$$\inf_{t>0}\mathbb{E}\left[\psi(tZ)\right]\ge\mathbb{E}\left[\mathbf{1}_{(0,\infty)}(Z)\right].$$

Consequently, the constraint

$$\inf_{t>0}\mathbb{E}\left[\psi(tZ_x)\right]\le\alpha \tag{6.20}$$

is a *conservative* approximation of the chance constraint (6.18) in the sense that the feasible set defined by (6.20) is contained in the feasible set defined by (6.18).

Of course, the smaller the function $\psi(\cdot)$ is the better this approximation will be. From this point of view the best choice of $\psi(\cdot)$ is to take piecewise linear function $\psi(z):=$

[39] Average Value-at-Risk is often called Conditional Value-at-Risk. We adopt here the term "Average" rather than "Conditional" Value-at-Risk in order to avoid awkward notation and terminology while discussing later conditional risk mappings.

[40] Recall that $\mathbf{1}_{(0,\infty)}(z)=0$ if $z\le 0$ and $\mathbf{1}_{(0,\infty)}(z)=1$ if $z>0$.

$[1+\gamma z]_+$ for some $\gamma > 0$. Since constraint (6.20) is invariant with respect to scale change of $\psi(\gamma z)$ to $\psi(z)$, we have that $\psi(z) := [1+z]_+$ gives the best choice of such a function. For this choice of function $\psi(\cdot)$, we have that constraint (6.20) is equivalent to

$$\inf_{t>0}\{t\mathbb{E}[t^{-1}+Z]_+ - \alpha\} \le 0,$$

or equivalently

$$\inf_{t>0}\left\{\alpha^{-1}\mathbb{E}[Z+t^{-1}]_+ - t^{-1}\right\} \le 0.$$

Now replacing t with $-t^{-1}$ we get the form

$$\inf_{t<0}\left\{t+\alpha^{-1}\mathbb{E}[Z-t]_+\right\} \le 0. \tag{6.21}$$

The quantity

$$\mathsf{AV@R}_\alpha(Z) := \inf_{t\in\mathbb{R}}\left\{t+\alpha^{-1}\mathbb{E}[Z-t]_+\right\} \tag{6.22}$$

is called the *Average Value-at-Risk*[41] of Z (at level α). Note that $\mathsf{AV@R}_\alpha(Z)$ is well defined and finite valued for every $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$.

The function $\varphi(t) := t + \alpha^{-1}\mathbb{E}[Z-t]_+$ is convex. Its derivative at t is equal to $1 + \alpha^{-1}[H_Z(t)-1]$, provided that the cdf $H_Z(\cdot)$ is continuous at t. If $H_Z(\cdot)$ is discontinuous at t, then the respective right- and left-side derivatives of $\varphi(\cdot)$ are given by the same formula with $H_Z(t)$ understood as the corresponding right- and left-side limits. Therefore the minimum of $\varphi(t)$, over $t \in \mathbb{R}$, is attained on the interval $[t^*, t^{**}]$, where

$$t^* := \inf\{z : H_Z(z) \ge 1-\alpha\} \ \text{ and } \ t^{**} := \sup\{z : H_Z(z) \le 1-\alpha\} \tag{6.23}$$

are the respective left- and right-side quantiles. Recall that the left-side quantile $t^* = \mathsf{V@R}_\alpha(Z)$.

Since the minimum of $\varphi(t)$ is attained at $t^* = \mathsf{V@R}_\alpha(Z)$, we have that $\mathsf{AV@R}_\alpha(Z)$ is bigger than $\mathsf{V@R}_\alpha(Z)$ by the nonnegative amount of $\alpha^{-1}\mathbb{E}[Z-t^*]_+$. Therefore

$$\inf_{t\in\mathbb{R}}\left\{t+\alpha^{-1}\mathbb{E}[Z-t]_+\right\} \le 0 \ \text{ implies that } t^* \le 0,$$

and hence constraint (6.21) is equivalent to $\mathsf{AV@R}_\alpha(Z) \le 0$. Therefore, the constraint

$$\mathsf{AV@R}_\alpha[Z_x] \le 0 \tag{6.24}$$

is equivalent to the constraint (6.21) and gives a conservative approximation[42] of the chance constraint (6.18).

The function $\rho(Z) := \mathsf{AV@R}_\alpha(Z)$, defined on a space of random variables, is *convex*, i.e., if Z and Z' are two random variables and $t \in [0,1]$, then

$$\rho\left(tZ+(1-t)Z'\right) \le t\rho(Z) + (1-t)\rho(Z').$$

[41] In some publications the concept of Average Value-at-Risk is called Conditional Value-at-Risk and is denoted $\mathsf{CV@R}_\alpha$.

[42] It is easy to see that for any $\tau \in \mathbb{R}$,

$$\mathsf{AV@R}_\alpha(Z+\tau) = \mathsf{AV@R}_\alpha(Z) + \tau. \tag{6.25}$$

Consequently, the constraint $\mathsf{AV@R}_\alpha[Z_x] \le \tau$ gives a conservative approximation of the chance constraint $\mathsf{V@R}_\alpha[Z_x] \le \tau$.

This follows from the fact that the function $t + \alpha^{-1}\mathbb{E}[Z - t]_+$ is convex jointly in t and Z. Also $\rho(\cdot)$ is monotone, i.e., if Z and Z' are two random variables such that with probability one $Z \geq Z'$, then $\rho(Z) \geq \rho(Z')$. It follows that if $G(\cdot, \xi)$ is convex for a.e. $\xi \in \Xi$, then the function $\rho[G(\cdot, \xi)]$ is also convex. Indeed, by convexity of $G(\cdot, \xi)$ and monotonicity of $\rho(\cdot)$, we have for any $t \in [0, 1]$ that

$$\rho[G(tZ + (1-t)Z', \xi)] \leq \rho[tG(Z, \xi) + (1-t)G(Z', \xi)]$$

and hence by convexity of $\rho(\cdot)$ that

$$\rho[G(tZ + (1-t)Z', \xi)] \leq t\rho[G(Z, \xi)] + (1-t)\rho[G(Z', \xi)].$$

Consequently, (6.24) is a *convex* conservative approximation of the chance constraint (6.18). Moreover, from the considered point of view, (6.24) is the best convex conservative approximation of the chance constraint (6.18).

We can now relate the concept of Average Value-at-Risk to mean deviations from quantiles. Recall that (see (6.14))

$$\mathfrak{q}_\alpha[Z] := \mathbb{E}\Big[\max\Big\{(1-\alpha)\Big(H_Z^{-1}(\alpha) - Z\Big), \alpha\Big(Z - H_Z^{-1}(\alpha)\Big)\Big\}\Big].$$

Theorem 6.2. *Let $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$ and $H(z)$ be its* cdf. *Then the following identities hold true:*

$$\mathsf{AV@R}_\alpha(Z) = \frac{1}{\alpha}\int_{1-\alpha}^{1} \mathsf{V@R}_{1-\tau}(Z)\,d\tau = \mathbb{E}[Z] + \frac{1}{\alpha}\mathfrak{q}_{1-\alpha}[Z]. \tag{6.26}$$

Moreover, if $H(z)$ is continuous at $z = \mathsf{V@R}_\alpha(Z)$, then

$$\mathsf{AV@R}_\alpha(Z) = \frac{1}{\alpha}\int_{\mathsf{V@R}_\alpha(Z)}^{+\infty} z\,dH(z) = \mathbb{E}\left[Z \middle| Z \geq \mathsf{V@R}_\alpha(Z)\right]. \tag{6.27}$$

Proof. As discussed earlier, the minimum in (6.22) is attained at $t^* = H^{-1}(1-\alpha) = \mathsf{V@R}_\alpha(Z)$. Therefore

$$\mathsf{AV@R}_\alpha(Z) = t^* + \alpha^{-1}\mathbb{E}[Z - t^*]_+ = t^* + \alpha^{-1}\int_{t^*}^{+\infty}(z - t^*)dH(z).$$

Moreover,

$$\int_{t^*}^{+\infty} dH(z) = \Pr(Z \geq t^*) = 1 - \Pr(Z \leq t^*) = \alpha,$$

provided that $\Pr(Z = t^*) = 0$, i.e., that $H(z)$ is continuous at $z = \mathsf{V@R}_\alpha(Z)$. This shows the first equality in (6.27), and then the second equality in (6.27) follows provided that $\Pr(Z = t^*) = 0$.

The first equality in (6.26) follows from the first equality in (6.27) by the substitution $\tau = H(z)$. Finally, we have

$$\begin{aligned}\mathsf{AV@R}_\alpha(Z) &= t^* + \alpha^{-1}\mathbb{E}[Z - t^*]_+ = \mathbb{E}[Z] + \mathbb{E}\left\{-Z + t^* + \alpha^{-1}[Z - t^*]_+\right\}\\ &= \mathbb{E}[Z] + \mathbb{E}\left[\max\left\{\alpha^{-1}(1-\alpha)(Z - t^*), t^* - Z\right\}\right].\end{aligned}$$

This proves the last equality in (6.26). $\square$

The first equation in (6.26) motivates the term *Average Value-at-Risk*. The last equation in (6.27) explains the origin of the alternative term *Conditional Value-at-Risk*.

Theorem 6.2 allows us to show an important relation between the absolute semideviation $\sigma_1^+[Z]$ and the mean deviation from quantile $\mathfrak{q}_\alpha[Z]$.

Corollary 6.3. *For every $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$ we have*

$$\sigma_1^+[Z] = \max_{\alpha\in[0,1]} \mathfrak{q}_\alpha[Z] = \min_{t\in\mathbb{R}} \max_{\alpha\in[0,1]} \mathbb{E}\{(1-\alpha)[t-Z]_+ + \alpha[Z-t]_+\}. \tag{6.28}$$

Proof. From (6.26) we get

$$\mathfrak{q}_{1-\alpha}[Z] = \int_{1-\alpha}^1 H_Z^{-1}(\tau)\,d\tau - \alpha\mathbb{E}[Z].$$

The right derivative of the right-hand side with respect to α equals $H_Z^{-1}(1-\alpha) - \mathbb{E}[Z]$. As it is nonincreasing, the function $\alpha \mapsto \mathfrak{q}_{1-\alpha}[Z]$ is concave. Moreover, its maximum is achieved at α^* for which $\mathbb{E}[Z]$ is the $(1-\alpha^*)$-quantile of Z. Substituting the minimizer $t^* = \mathbb{E}[Z]$ into (6.22) we conclude that

$$\mathsf{AV@R}_{\alpha^*}(Z) = \mathbb{E}[Z] + \frac{1}{\alpha^*}\sigma_1^+[Z].$$

Comparison with (6.26) yields the first equality in (6.28). To prove the second equality we recall relation (6.15) and note that

$$\max\big((1-\alpha)(t-Z), \alpha(Z-t)\big) = (1-\alpha)[t-Z]_+ + \alpha[Z-t]_+.$$

Thus

$$\sigma_1^+[Z] = \max_{\alpha\in[0,1]} \min_{t\in\mathbb{R}} \mathbb{E}\{(1-\alpha)[t-Z]_+ + \alpha[Z-t]_+\}.$$

As the function under the max-min operation is linear with respect to $\alpha \in [0,1]$ and convex with respect to t, the max and min operations can be exchanged. This proves the second equality in (6.28). $\square$

It also follows from (6.26) that the minimization problem (6.16) can be equivalently written as follows:

$$\begin{aligned}\min_{x\in X} \mathbb{E}[Z_x] + c\mathfrak{q}_{1-\alpha}[Z_x] &= \min_{x\in X} (1-c\alpha)\mathbb{E}[Z_x] + c\alpha\,\mathsf{AV@R}_\alpha[Z_x] \\ &= \min_{x\in X, t\in\mathbb{R}} \mathbb{E}\big[(1-c\alpha)Z_x + c\big(\alpha t + [Z_x - t]_+\big)\big].\end{aligned} \tag{6.29}$$

Both x and t are variables in this problem. We conclude that for this specific mean–risk model, an equivalent expected value formulation has been found. If $c \in [0, \alpha^{-1}]$ and the function $x \mapsto Z_x$ is convex, problem (6.29) is convex.

The maximization problem (6.17) can be equivalently written as follows:

$$\max_{x\in X} \mathbb{E}[Z_x] - c\mathfrak{q}_\alpha[Z_x] = -\min_{x\in X} \mathbb{E}[-Z_x] + c\mathfrak{q}_{1-\alpha}[-Z_x] \tag{6.30}$$

$$= -\min_{x\in X, t\in\mathbb{R}} \mathbb{E}\big[-(1-c\alpha)Z_x + c\big(\alpha t + [-Z_x - t]_+\big)\big]$$

$$= \max_{x\in X, t\in\mathbb{R}} \mathbb{E}\big[(1-c\alpha)Z_x + c\big(\alpha t - [t - Z_x]_+\big)\big]. \tag{6.31}$$

In the last problem we replaced t by$-t$ to stress the similarity with (6.29). Again, if $c \in [0, \alpha^{-1}]$ and the function $x \mapsto Z_x$ is convex, problem (6.30) is convex.

6.3 Coherent Risk Measures

Let $(\Omega, \mathcal{F})$ be a sample space, equipped with the sigma algebra $\mathcal{F}$, on which considered uncertain outcomes (random functions $Z = Z(\omega)$) are defined. By a *risk measure* we understand a function $\rho(Z)$ which maps Z into the extended real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$. In order to make this concept precise we need to define a space $\mathcal{Z}$ of allowable random functions $Z(\omega)$ for which $\rho(Z)$ is defined. It seems that a natural choice of $\mathcal{Z}$ will be the space of all $\mathcal{F}$-measurable functions $Z : \Omega \to \mathbb{R}$. However, typically, this space is too large for development of a meaningful theory. Unless stated otherwise, we deal in this chapter with spaces $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$, where $p \in [1, +\infty)$. (See section 7.3 for an introduction of these spaces.) By assuming that $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$, we assume that random variable $Z(\omega)$ has a finite pth order moment with respect to the reference probability measure P. Also, by considering function ρ to be defined on the space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$, it is implicitly assumed that actually ρ is defined on classes of functions which can differ on sets of P-measure zero, i.e., $\rho(Z) = \rho(Z')$ if $P\{\omega : Z(\omega) \neq Z'(\omega)\} = 0$.

We assume throughout this chapter that risk measures $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ are *proper*. That is, $\rho(Z) > -\infty$ for all $Z \in \mathcal{Z}$ and the domain

$$\operatorname{dom}(\rho) := \{Z \in \mathcal{Z} : \rho(Z) < +\infty\}$$

is nonempty. We consider the following axioms associated with a risk measure ρ. For $Z, Z' \in \mathcal{Z}$ we denote by $Z \succeq Z'$ the pointwise partial order,[43] meaning $Z(\omega) \geq Z'(\omega)$ for a.e. $\omega \in \Omega$. We also assume that the smaller the realizations of Z, the better; for example, Z may represent a random cost.

(R1) Convexity:

$$\rho(tZ + (1-t)Z') \leq t\rho(Z) + (1-t)\rho(Z')$$

for all $Z, Z' \in \mathcal{Z}$ and all $t \in [0, 1]$.

(R2) Monotonicity: If $Z, Z' \in \mathcal{Z}$ and $Z \succeq Z'$, then $\rho(Z) \geq \rho(Z')$.

(R3) Translation equivariance: If $a \in \mathbb{R}$ and $Z \in \mathcal{Z}$, then $\rho(Z + a) = \rho(Z) + a$.

(R4) Positive homogeneity: If $t > 0$ and $Z \in \mathcal{Z}$, then $\rho(tZ) = t\rho(Z)$.

It is said that a risk measure ρ is *coherent* if it satisfies the above conditions (R1)–(R4). An example of a coherent risk measure is the Average Value-at-Risk $\rho(Z) := \mathsf{AV@R}_\alpha(Z)$. (Further examples of risk measures will be discussed in section 6.3.2.) It is natural to assume in this example that Z has a finite first order moment, i.e., to use $\mathcal{Z} := \mathcal{L}_1(\Omega, \mathcal{F}, P)$. For such space $\mathcal{Z}$ in this example, $\rho(Z)$ is finite (real valued) for all $Z \in \mathcal{Z}$.

[43]This partial order corresponds to the cone $\mathcal{C} := \mathcal{L}_p^+(\Omega, \mathcal{F}, P)$. See the discussion of section 7.3, page 404, following (7.245).

If the random outcome represents a reward, i.e., larger realizations of Z are preferred, we can define a risk measure $\varrho(Z) = \rho(-Z)$, where ρ satisfies axioms (R1)–(R4). In this case, the function ϱ also satisfies (R1) and (R4). The axioms (R2) and (R3) change to

(R2a) Monotonicity: If $Z, Z' \in \mathcal{Z}$ and $Z \succeq Z'$, then $\varrho(Z) \le \varrho(Z')$.

(R3a) Translation equivariance: If $a \in \mathbb{R}$ and $Z \in \mathcal{Z}$, then $\varrho(Z + a) = \varrho(Z) - a$.

All our considerations regarding risk measures satisfying (R1)–(R4) have their obvious counterparts for risk measures satisfying (R1), (R2a), (R3a), and (R4).

With each space $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ is associated its dual space $\mathcal{Z}^* := \mathcal{L}_q(\Omega, \mathcal{F}, P)$, where $q \in (1, +\infty]$ is such that $1/p + 1/q = 1$. For $Z \in \mathcal{Z}$ and $\zeta \in \mathcal{Z}^*$ their scalar product is defined as

$$\langle \zeta, Z \rangle := \int_\Omega \zeta(\omega) Z(\omega) dP(\omega). \tag{6.32}$$

Recall that the conjugate function $\rho^* : \mathcal{Z}^* \to \overline{\mathbb{R}}$ of a risk measure ρ is defined as

$$\rho^*(\zeta) := \sup_{Z \in \mathcal{Z}} \big\{ \langle \zeta, Z \rangle - \rho(Z) \big\} \tag{6.33}$$

and the conjugate of ρ^* (the biconjugate function) as

$$\rho^{**}(Z) := \sup_{\zeta \in \mathcal{Z}^*} \big\{ \langle \zeta, Z \rangle - \rho^*(\zeta) \big\}. \tag{6.34}$$

By the Fenchel–Moreau theorem (Theorem 7.71) we have that if $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is convex, proper and lower semicontinuous, then $\rho^{**} = \rho$, i.e., $\rho(\cdot)$ has the representation

$$\rho(Z) = \sup_{\zeta \in \mathcal{Z}^*} \big\{ \langle \zeta, Z \rangle - \rho^*(\zeta) \big\}, \quad \forall Z \in \mathcal{Z}. \tag{6.35}$$

Conversely, if the representation (6.35) holds for some proper function $\rho^* : \mathcal{Z}^* \to \overline{\mathbb{R}}$, then ρ is convex, proper, and lower semicontinuous. Note that if ρ is convex, proper, and lower semicontinuous, then its conjugate function ρ^* is also proper. Clearly, we can write the representation (6.35) in the following equivalent form:

$$\rho(Z) = \sup_{\zeta \in \mathfrak{A}} \big\{ \langle \zeta, Z \rangle - \rho^*(\zeta) \big\}, \quad \forall Z \in \mathcal{Z}, \tag{6.36}$$

where $\mathfrak{A} := \mathrm{dom}(\rho^*)$ is the domain of the conjugate function ρ^*.

The following basic duality result for convex risk measures is a direct consequence of the Fenchel–Moreau theorem.

Theorem 6.4. *Suppose that $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is convex, proper, and lower semicontinuous. Then the representation* (6.36) *holds with $\mathfrak{A} := \mathrm{dom}(\rho^*)$. Moreover, we have that:* (i) *condition* (R2) *holds iff every $\zeta \in \mathfrak{A}$ is nonnegative, i.e., $\zeta(\omega) \ge 0$ for a.e. $\omega \in \Omega$;* (ii) *condition* (R3) *holds iff $\int_\Omega \zeta dP = 1$ for every $\zeta \in \mathfrak{A}$; and* (iii) *condition* (R4) *holds iff $\rho(\cdot)$ is the support function of the set $\mathfrak{A}$, i.e., can be represented in the form*

$$\rho(Z) = \sup_{\zeta \in \mathfrak{A}} \langle \zeta, Z \rangle, \quad \forall Z \in \mathcal{Z}. \tag{6.37}$$

Proof. If $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is convex, proper, and lower semicontinuous, then representation (6.36) is valid by virtue of the Fenchel–Moreau theorem (Theorem 7.71).

Now suppose that assumption (R2) holds. It follows that $\rho^*(\zeta) = +\infty$ for every $\zeta \in \mathcal{Z}^*$ which is not nonnegative. Indeed, if $\zeta \in \mathcal{Z}^*$ is not nonnegative, then there exists a set $\Delta \in \mathcal{F}$ of positive measure such that $\zeta(\omega) < 0$ for all $\omega \in \Delta$. Consequently, for $\bar{Z} := \mathbf{1}_\Delta$ we have that $\langle \zeta, \bar{Z} \rangle < 0$. Take any Z in the domain of ρ, i.e., such that $\rho(Z)$ is finite, and consider $Z_t := Z - t\bar{Z}$. Then for $t \geq 0$, we have that $Z \succeq Z_t$, and assumption (R2) implies that $\rho(Z) \geq \rho(Z_t)$. Consequently,

$$\rho^*(\zeta) \geq \sup_{t \in \mathbb{R}_+} \big\{ \langle \zeta, Z_t \rangle - \rho(Z_t) \big\} \geq \sup_{t \in \mathbb{R}_+} \big\{ \langle \zeta, Z \rangle - t \langle \zeta, \bar{Z} \rangle - \rho(Z) \big\} = +\infty.$$

Conversely, suppose that every $\zeta \in \mathfrak{A}$ is nonnegative. Then for every $\zeta \in \mathfrak{A}$ and $Z \succeq Z'$, we have that $\langle \zeta, Z' \rangle \geq \langle \zeta, Z \rangle$. By (6.36), this implies that if $Z \succeq Z'$, then $\rho(Z) \geq \rho(Z')$. This completes the proof of assertion (i).

Suppose that assumption (R3) holds. Then for every $Z \in \operatorname{dom}(\rho)$ we have

$$\rho^*(\zeta) \geq \sup_{a \in \mathbb{R}} \big\{ \langle \zeta, Z + a \rangle - \rho(Z + a) \big\} = \sup_{a \in \mathbb{R}} \left\{ a \int_\Omega \zeta \, dP - a + \langle \zeta, Z \rangle - \rho(Z) \right\}.$$

It follows that $\rho^*(\zeta) = +\infty$ for any $\zeta \in \mathcal{Z}^*$ such that $\int_\Omega \zeta \, dP \neq 1$. Conversely, if $\int_\Omega \zeta \, dP = 1$, then $\langle \zeta, Z + a \rangle = \langle \zeta, Z \rangle + a$, and hence condition (R3) follows by (6.36). This completes the proof of (ii).

Clearly, if (6.37) holds, then ρ is positively homogeneous. Conversely, if ρ is positively homogeneous, then its conjugate function is the indicator function of a convex subset of $\mathcal{Z}^*$. Consequently, the representation (6.37) follows by (6.36). ∎

It follows from the above theorem that if ρ is a risk measure satisfying conditions (R1)–(R3) and is proper and lower semicontinuous, then the representation (6.36) holds with $\mathfrak{A}$ being a subset of the set of probability density functions,

$$\mathfrak{P} := \left\{ \zeta \in \mathcal{Z}^* : \int_\Omega \zeta(\omega) dP(\omega) = 1, \ \zeta \succeq 0 \right\}. \tag{6.38}$$

If, moreover, ρ is positively homogeneous (i.e., condition (R4) holds), then its conjugate ρ^* is the indicator function of a convex set $\mathfrak{A} \subset \mathcal{Z}^*$, and $\mathfrak{A}$ is equal to the subdifferential $\partial\rho(0)$ of ρ at $0 \in \mathcal{Z}$. Furthermore, $\rho(0) = 0$ and hence by the definition of $\partial\rho(0)$ we have that

$$\mathfrak{A} = \{ \zeta \in \mathfrak{P} : \langle \zeta, Z \rangle \leq \rho(Z), \quad \forall\, Z \in \mathcal{Z} \}. \tag{6.39}$$

The set $\mathfrak{A}$ is weakly* closed. Recall that if the space $\mathcal{Z}$, and hence $\mathcal{Z}^*$, is reflexive, then a convex subset of $\mathcal{Z}^*$ is closed in the weak* topology of $\mathcal{Z}^*$ iff it is closed in the strong (norm) topology of $\mathcal{Z}^*$. If ρ is positively homogeneous and continuous, then $\mathfrak{A} = \partial\rho(0)$ is a *bounded* (and weakly* compact) subset of $\mathcal{Z}^*$ (see Proposition 7.74).

We have that if ρ is a *coherent* risk measure, then the corresponding set $\mathfrak{A}$ is a set of probability density functions. Consequently, for any $\zeta \in \mathfrak{A}$ we can view $\langle \zeta, Z \rangle$ as the expectation $\mathbb{E}_\zeta[Z]$ taken with respect to the probability measure $\zeta \, dP$, defined by the density ζ. Consequently representation (6.37) can be written in the form

$$\rho(Z) = \sup_{\zeta \in \mathfrak{A}} \mathbb{E}_\zeta[Z], \quad \forall Z \in \mathcal{Z}. \tag{6.40}$$

Definition of a risk measure ρ depends on a particular choice of the corresponding space $\mathcal{Z}$. In many cases there is a natural choice of $\mathcal{Z}$ which ensures that $\rho(Z)$ is finite valued for all $Z \in \mathcal{Z}$. We shall see such examples in section 6.3.2. By Theorem 7.79 we have the following result, which shows that for real valued convex and monotone risk measures, the assumption of lower semicontinuity in Theorem 6.4 holds automatically.

Proposition 6.5. *Let $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ with $p \in [1, +\infty]$ and $\rho : \mathcal{Z} \to \mathbb{R}$ be a real valued risk measure satisfying conditions* (R1) *and* (R2). *Then ρ is continuous and subdifferentiable on $\mathcal{Z}$.*

Theorem 6.4 together with Proposition 6.5 imply the following basic duality result.

Theorem 6.6. *Let $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$, where $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ with $p \in [1, +\infty)$. Then ρ is a real valued coherent risk measure iff there exists a convex bounded and weakly* closed set $\mathfrak{A} \subset \mathfrak{P}$ such that the representation* (6.37) *holds.*

Proof. If $\rho : \mathcal{Z} \to \mathbb{R}$ is a real valued coherent risk measure, then by Proposition 6.5 it is continuous, and hence by Theorem 6.4 the representation (6.37) holds with $\mathfrak{A} = \partial\rho(0)$. Moreover, the subdifferential of a convex continuous function is bounded and weakly* closed (and hence is weakly* compact).

Conversely, if the representation (6.37) holds with the set $\mathfrak{A}$ being a convex subset of $\mathfrak{P}$ and weakly* compact, then ρ is real valued and satisfies conditions (R1)–(R4). □

The following result shows that if a risk measure satisfies conditions (R1)–(R3), then either it is finite valued and continuous on $\mathcal{Z}$ or it takes value $+\infty$ on a dense subset of $\mathcal{Z}$.

Proposition 6.7. *Let $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$, with $p \in [1, +\infty)$, and $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ be a proper risk measure satisfying conditions* (R1), (R2) *and* (R3). *Suppose that the domain of ρ has a nonempty interior. Then ρ is finite valued and continuous on $\mathcal{Z}$.*

Proof. Consider the level sets of ρ:

$$\mathcal{A}_c := \{Z \in \mathcal{Z} : \rho(Z) \le c\}.$$

We have that $\cup_{c\in\mathbb{N}}\mathcal{A}_c = \mathrm{dom}(\rho)$. Since $\mathrm{dom}(\rho)$ has a nonempty interior, it follows by Baire's lemma that for some $c \in \mathbb{R}$ the set $\mathcal{A}_c$ has a nonempty interior. Because of condition (R3) we have that $Z \in \mathcal{A}_0$ iff $Z + c \in \mathcal{A}_c$, i.e., $\mathcal{A}_c = \mathcal{A}_0 + c$ (here c denotes the constant function $Z(\cdot) = c$). Therefore $\mathcal{A}_0$ has a nonempty interior. That is, there exist $Z_0 \in \mathcal{Z}$ and $r > 0$ such that $B(Z_0, r) \subset \mathcal{A}_0$, where

$$B(Z_0, r) := \{Z \in \mathcal{Z} : \|Z - Z_0\| \le r\}.$$

By changing variables $Z \mapsto Z - Z_0$, we can assume without loss of generality that $Z_0 = 0$, i.e., $B(0, r) \subset \mathcal{A}_0$.

Consider a point $Z \in \mathcal{Z}$. For $c \in \mathbb{R}$ we have that $Z = Z_c^- + Z_c^+$, where $Z_c^-(\cdot) := \min\{Z(\cdot), c\}$ and $Z_c^+(\cdot) := [Z(\cdot) - c]_+$. Note that for c large enough, the norm of Z_c^+ can be made arbitrarily small. Therefore we can choose c such that $\|Z_c^+\| < r$. Since $\mathcal{A}_c = \mathcal{A}_0 + c$, we have that $B(c, r) \subset \mathcal{A}_c$. Consequently, $c + Z_c^+ \in \mathcal{A}_c$. It follows by the monotonicity condition (R2) that $\rho(Z) \le \rho(c + Z_c^+) \le c$, and hence $\rho(Z)$ is finite. That is, we showed that $\rho(\cdot)$ is finite valued on $\mathcal{Z}$. Continuity of $\rho(\cdot)$ follows by Proposition 6.5. $\square$

It is not difficult to show (we leave this as an exercise) that for $\mathcal{Z} := \mathcal{L}_\infty(\Omega, \mathcal{F}, P)$ any risk measure $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ satisfying conditions (R1)–(R3), and having a finite value in at least one point of $\mathcal{Z}$, is finite valued and hence is continuous by Proposition 6.5.

Of course, the analysis simplifies considerably if the space Ω is finite, say, $\Omega := \{\omega_1, \dots, \omega_K\}$ equipped with sigma algebra of all subsets of Ω and respective (positive) probabilities $p_1, \dots, p_K$. Then every function $Z : \Omega \to \mathbb{R}$ is measurable and the space $\mathcal{Z}$ of all such functions can be identified with $\mathbb{R}^K$ by identifying $Z \in \mathcal{Z}$ with the vector $(Z(\omega_1), \dots, Z(\omega_K)) \in \mathbb{R}^K$. The dual of the space $\mathbb{R}^K$ can be identified with itself by using the standard scalar product in $\mathbb{R}^K$, and the set $\mathfrak{P}$ becomes

$$\mathfrak{P} = \left\{ \zeta \in \mathbb{R}^K : \sum_{k=1}^K p_k \zeta_k = 1,\ \zeta \ge 0 \right\}. \tag{6.41}$$

The above set $\mathfrak{P}$ forms a convex bounded subset of $\mathbb{R}^K$, and hence the set $\mathfrak{A}$ is also bounded.

6.3.1 Differentiability Properties of Risk Measures

Let $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ be a convex proper lower semicontinuous risk measure. By convexity and lower semicontinuity of ρ we have that $\rho^{**} = \rho$ and hence by Proposition 7.73 that

$$\partial\rho(Z) = \arg\max_{\zeta\in\mathfrak{A}} \left\{ \langle \zeta, Z\rangle - \rho^*(\zeta) \right\}, \tag{6.42}$$

provided that $\rho(Z)$ is finite. If, moreover, ρ is positively homogeneous, then $\mathfrak{A} = \partial\rho(0)$ and

$$\partial\rho(Z) = \arg\max_{\zeta\in\mathfrak{A}} \langle \zeta, Z\rangle. \tag{6.43}$$

As we know, conditions (R1)–(R3) imply that $\mathfrak{A}$ is a subset of the set $\mathfrak{P}$ of probability density functions. Consequently, under conditions (R1)–(R3), $\partial\rho(Z)$ is a subset of $\mathfrak{P}$ as well.

We also have that if ρ is finite valued and continuous at Z, then $\partial\rho(Z)$ is a nonempty bounded and weakly* compact subset of $\mathcal{Z}^*$, ρ is Hadamard directionally differentiable and subdifferentiable at Z, and

$$\rho'(Z, H) = \sup_{\zeta\in\partial\rho(Z)} \langle \zeta, H\rangle, \quad \forall H \in \mathcal{Z}. \tag{6.44}$$

In particular, if ρ is continuous at Z and $\partial\rho(Z)$ is a singleton, i.e., $\partial\rho(Z)$ consists of unique point denoted $\nabla\rho(Z)$, then ρ is Hadamard differentiable at Z and

$$\rho'(Z, \cdot) = \langle \nabla\rho(Z), \cdot\rangle. \tag{6.45}$$

We often have to deal with composite functions $\rho \circ F : \mathbb{R}^n \to \overline{\mathbb{R}}$, where $F : \mathbb{R}^n \to \mathcal{Z}$ is a mapping. We write $f(x, \omega)$, or $f_\omega(x)$, for $[F(x)](\omega)$, and view $f(x, \omega)$ as a random function defined on the measurable space $(\Omega, \mathcal{F})$. We say that the mapping F is *convex* if the function $f(\cdot, \omega)$ is convex for every $\omega \in \Omega$.

Proposition 6.8. *If the mapping $F : \mathbb{R}^n \to \mathcal{Z}$ is convex and $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ satisfies conditions* (R1)–(R2)*, then the composite function $\phi(\cdot) := \rho(F(\cdot))$ is convex.*

Proof. For any $x, x' \in \mathbb{R}^n$ and $t \in [0, 1]$, we have by convexity of $F(\cdot)$ and monotonicity of $\rho(\cdot)$ that

$$\rho(F(tx + (1-t)x')) \le \rho(tF(x) + (1-t)F(x')).$$

Hence convexity of $\rho(\cdot)$ implies that

$$\rho(F(tx + (1-t)x')) \le t\rho(F(x)) + (1-t)\rho(F(x')).$$

This proves convexity of $\rho(F(\cdot))$. $\square$

It should be noted that the monotonicity condition (R2) was essential in the above derivation of convexity of the composite function.

Let us discuss differentiability properties of the composite function $\phi = \rho \circ F$ at a point $\bar{x} \in \mathbb{R}^n$. As before, we assume that $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$. The mapping $F : \mathbb{R}^n \to \mathcal{Z}$ maps a point $x \in \mathbb{R}^n$ into a real valued function (or rather a class of functions which may differ on sets of P-measure zero) $[F(x)](\cdot)$ on Ω, also denoted $f(x, \cdot)$, which is an element of $\mathcal{L}_p(\Omega, \mathcal{F}, P)$. If F is convex, then $f(\cdot, \omega)$ is convex real valued and hence is continuous and has (finite valued) directional derivatives at $\bar{x}$, denoted $f'_\omega(\bar{x}, h)$. These properties are inherited by the mapping F.

Lemma 6.9. *Let $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and $F : \mathbb{R}^n \to \mathcal{Z}$ be a convex mapping. Then F is continuous and directionally differentiable, and*

$$[F'(\bar{x}, h)](\omega) = f'_\omega(\bar{x}, h), \quad \omega \in \Omega, \; h \in \mathbb{R}^n. \tag{6.46}$$

Proof. In order to show continuity of F we need to verify that, for an arbitrary point $\bar{x} \in \mathbb{R}^n$, $\|F(x) - F(\bar{x})\|_p$ tends to zero as $x \to \bar{x}$. By the Lebesgue dominated convergence theorem and continuity of $f(\cdot, \omega)$ we can write that

$$\lim_{x \to \bar{x}} \int_\Omega \left| f(x, \omega) - f(\bar{x}, \omega) \right|^p dP(\omega) = \int_\Omega \lim_{x \to \bar{x}} \left| f(x, \omega) - f(\bar{x}, \omega) \right|^p dP(\omega) = 0, \tag{6.47}$$

provided that there exists a neighborhood $U \subset \mathbb{R}^n$ of $\bar{x}$ such that the family $\{|f(x, \omega) - f(\bar{x}, \omega)|^p\}_{x \in U}$ is dominated by a P-integrable function, or equivalently that $\{|f(x, \omega) - f(\bar{x}, \omega)|\}_{x \in U}$ is dominated by a function from the space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$. Since $f(\bar{x}, \cdot)$ belongs to $\mathcal{L}_p(\Omega, \mathcal{F}, P)$, it suffices to verify this dominance condition for $\{|f(x, \omega)|\}_{x \in U}$. Now let $x_1, \dots, x_{n+1} \in \mathbb{R}^n$ be such points that the set $U := \text{conv}\{x_1, \dots, x_{n+1}\}$ forms a neighborhood of the point $\bar{x}$, and let $g(\omega) := \max\{f(x_1, \omega), \dots, f(x_{n+1}, \omega)\}$. By convexity of $f(\cdot, \omega)$ we have that $f(x, \cdot) \le g(\cdot)$ for all $x \in U$. Also since every $f(x_i, \dots)$, $i = 1, \dots, n+1$, is an element of $\mathcal{L}_p(\Omega, \mathcal{F}, P)$, we have that $g \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$ as well. That is, $g(\omega)$ gives an upper bound for $\{|f(x,\omega)|\}_{x\in U}$. Also by convexity of $f(\cdot,\omega)$ we have that

$$f(x,\omega) \geq 2f(\bar{x},\omega) - f(2\bar{x}-x,\omega).$$

By shrinking the neighborhood U if necessary, we can assume that U is symmetrical around $\bar{x}$, i.e., if $x \in U$, then $2\bar{x}-x \in U$. Consequently, we have that $\tilde{g}(\omega) := 2f(\bar{x},\omega) - g(w)$ gives a lower bound for $\{|f(x,\omega)|\}_{x\in U}$, and $\tilde{g} \in \mathcal{L}_p(\Omega,\mathcal{F},P)$. This shows that the required dominance condition holds and hence F is continuous at $\bar{x}$ by (6.47).

Now for $h \in \mathbb{R}^n$ and $t > 0$ denote

$$R_t(\omega) := t^{-1}\left[f(\bar{x}+th,\omega) - f(\bar{x},\omega)\right] \text{ and } Z(\omega) := f'_\omega(\bar{x},h), \quad \omega \in \Omega.$$

Note that $f(\bar{x}+th,\cdot)$ and $f(\bar{x},\cdot)$ are elements of the space $\mathcal{L}_p(\Omega,\mathcal{F},P)$, and hence $R_t(\cdot)$ is also an element of $\mathcal{L}_p(\Omega,\mathcal{F},P)$ for any $t>0$. Since for a.e. $\omega\in\Omega$, $f(\cdot,\omega)$ is convex real valued, we have that $R_t(\omega)$ is monotonically nonincreasing and converges to $Z(\omega)$ as $t \downarrow 0$. Therefore, we have that $R_t(\cdot) \geq Z(\cdot)$ for any $t > 0$. Again by convexity of $f(\cdot,\omega)$, we have that for $t>0$,

$$Z(\cdot) \geq t^{-1}\left[f(\bar{x},\cdot) - f(\bar{x}-th,\cdot)\right].$$

We obtain that $Z(\cdot)$ is bounded from above and below by functions which are elements of the space $\mathcal{L}_p(\Omega,\mathcal{F},P)$ and hence $Z \in \mathcal{L}_p(\Omega,\mathcal{F},P)$ as well.

We have that $R_t(\cdot) - Z(\cdot)$, and hence $|R_t(\cdot)-Z(\cdot)|^p$, are monotonically decreasing to zero as $t\downarrow 0$ and for any $t>0$, $\mathbb{E}\left[|R_t - Z|^p\right]$ is finite. It follows by the monotone convergence theorem that $\mathbb{E}\left[|R_t - Z|^p\right]$ tends to zero as $t \downarrow 0$. That is, R_t converges to Z in the norm topology of $\mathcal{Z}$. Since $R_t = t^{-1}[F(\bar{x}+th) - F(\bar{x})]$, this shows that F is directionally differentiable at $\bar{x}$ and formula (6.46) follows. □

The following theorem can be viewed as an extension of Theorem 7.46, where a similar result is derived for $\rho(\cdot) := \mathbb{E}[\,\cdot\,]$.

Theorem 6.10. *Let $\mathcal{Z} := \mathcal{L}_p(\Omega,\mathcal{F},P)$ and $F:\mathbb{R}^n \to \mathcal{Z}$ be a convex mapping. Suppose that ρ is convex, finite valued, and continuous at $\bar{Z} := F(\bar{x})$. Then the composite function $\phi = \rho\circ F$ is directionally differentiable at $\bar{x}$, $\phi'(\bar{x},h)$ is finite valued for every $h\in\mathbb{R}^n$, and*

$$\phi'(\bar{x},h) = \sup_{\zeta\in\partial\rho(\bar{Z})} \int_\Omega f'_\omega(\bar{x},h)\zeta(\omega)dP(\omega). \tag{6.48}$$

Proof. Since ρ is continuous at $\bar{Z}$, it follows that ρ is subdifferentiable and Hadamard directionally differentiable at $\bar{Z}$ and formula (6.44) holds. Also by Lemma 6.9, mapping F is directionally differentiable. Consequently, we can apply the chain rule (see Proposition 7.58) to conclude that $\phi(\cdot)$ is directionally differentiable at $\bar{x}$, $\phi'(\bar{x},h)$ is finite valued and

$$\phi'(\bar{x},h) = \rho'(\bar{Z}, F'(\bar{x},h)). \tag{6.49}$$

Together with (6.44) and (6.46), the above formula (6.49) implies (6.48). □

It is also possible to write formula (6.48) in terms of the corresponding subdifferentials.

Theorem 6.11. *Let $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and $F : \mathbb{R}^n \to \mathcal{Z}$ be a convex mapping. Suppose that ρ satisfies conditions* (R1) *and* (R2) *and is finite valued and continuous at $\bar{Z} := F(\bar{x})$. Then the composite function $\phi = \rho \circ F$ is subdifferentiable at $\bar{x}$ and*

$$\partial\phi(\bar{x}) = \mathrm{cl}\left(\bigcup_{\zeta \in \partial\rho(\bar{Z})} \int_\Omega \partial f_\omega(\bar{x})\zeta(\omega)dP(\omega)\right). \tag{6.50}$$

Proof. Since, by Lemma 6.9, F is continuous at $\bar{x}$ and ρ is continuous at $F(\bar{x})$, we have that ϕ is continuous at $\bar{x}$, and hence $\phi(x)$ is finite valued for all x in a neighborhood of $\bar{x}$. Moreover, by Proposition 6.8, $\phi(\cdot)$ is convex and hence is continuous in a neighborhood of $\bar{x}$ and is subdifferentiable at $\bar{x}$. Also, formula (6.48) holds. It follows that $\phi'(\bar{x}, \cdot)$ is convex, continuous, and positively homogeneous, and

$$\phi'(\bar{x}, \cdot) = \sup_{\zeta \in \partial\rho(\bar{Z})} \eta_\zeta(\cdot), \tag{6.51}$$

where

$$\eta_\zeta(\cdot) := \int_\Omega f'_\omega(\bar{x}, \cdot)\zeta(\omega)dP(\omega). \tag{6.52}$$

Because of condition (R2), we have that every $\zeta \in \partial\rho(\bar{Z})$ is nonnegative. Consequently, the corresponding function η_ζ is convex continuous and positively homogeneous and hence is the support function of the set $\partial\eta_\zeta(0)$. The supremum of these functions, given by the right-hand side of (6.51), is the support function of the set $\cup_{\zeta\in\partial\rho(\bar{Z})}\partial\eta_\zeta(0)$. Applying Theorem 7.47 and using the fact that the subdifferential of $f'_\omega(\bar{x}, \cdot)$ at $0 \in \mathbb{R}^n$ coincides with $\partial f_\omega(\bar{x})$, we obtain

$$\partial\eta_\zeta(0) = \int_\Omega \partial f_\omega(\bar{x})\zeta(\omega)dP(\omega). \tag{6.53}$$

Since $\partial\rho(\bar{Z})$ is convex, it is straightforward to verify that the set $\cup_{\zeta\in\partial\rho(\bar{Z})}\partial\eta_\zeta(0)$ is also convex. Consequently it follows by (6.51) and (6.53) that the subdifferential of $\phi'(\bar{x}, \cdot)$ at $0 \in \mathbb{R}^n$ is equal to the topological closure of the set $\cup_{\zeta\in\partial\rho(\bar{Z})}\partial\eta_\zeta(0)$, i.e., is given by the right-hand side of (6.50). It remains to note that the subdifferential of $\phi'(\bar{x}, \cdot)$ at $0 \in \mathbb{R}^n$ coincides with $\partial\phi(\bar{x})$. $\square$

Under the assumptions of the above theorem, we have that the composite function ϕ is convex and is continuous (in fact, even Lipschitz continuous) in a neighborhood of $\bar{x}$. It follows that ϕ is differentiable[44] at $\bar{x}$ iff $\partial\phi(\bar{x})$ is a singleton. This leads to the following result, where for $\zeta \succeq 0$ we say that a property holds for ζ-a.e. $\omega \in \Omega$ if the set of points $A \in \mathcal{F}$ where it does not hold has ζdP measure zero, i.e., $\int_A \zeta(\omega)dP(\omega) = 0$. Of course, if $P(A) = 0$, then $\int_A \zeta(\omega)dP(\omega) = 0$. That is, if a property holds for a.e. $\omega \in \Omega$ with respect to P, then it holds for ζ-a.e. $\omega \in \Omega$.

[44]Note that since $\phi(\cdot)$ is Lipschitz continuous near $\bar{x}$, the notions of Gâteaux and Fréchet differentiability at $\bar{x}$ are equivalent here.

Corollary 6.12. *Let $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and $F : \mathbb{R}^n \to \mathcal{Z}$ be a convex mapping. Suppose that ρ satisfies conditions* (R1) *and* (R2) *and is finite valued and continuous at $\bar{Z} := F(\bar{x})$. Then the composite function $\phi = \rho \circ F$ is differentiable at $\bar{x}$ iff the following two properties hold:* (i) *for every $\zeta \in \partial\rho(\bar{Z})$ the function $f(\cdot, \omega)$ is differentiable at $\bar{x}$ for ζ-a.e. $\omega \in \Omega$, and* (ii) *$\int_\Omega \nabla f_\omega(\bar{x})\zeta(\omega) dP(\omega)$ is the same for every $\zeta \in \partial\rho(\bar{Z})$.*

In particular, if $\partial\rho(\bar{Z}) = \{\bar{\zeta}\}$ is a singleton, then ϕ is differentiable at $\bar{x}$ iff $f(\cdot, \omega)$ is differentiable at $\bar{x}$ for $\bar{\zeta}$-a.e. $\omega \in \Omega$, in which case

$$\nabla\phi(\bar{x}) = \int_\Omega \nabla f_\omega(\bar{x})\bar{\zeta}(\omega) dP(\omega). \tag{6.54}$$

Proof. By Theorem 6.11 we have that ϕ is differentiable at $\bar{x}$ iff the set on the right-hand side of (6.50) is a singleton. Clearly this set is a singleton iff the set $\int_\Omega \partial f_\omega(\bar{x})\zeta(\omega) dP(\omega)$ is a singleton and is the same for every $\zeta \in \partial\rho(\bar{Z})$. Since $\partial f_\omega(\bar{x})$ is a singleton iff $f_\omega(\cdot)$ is differentiable at $\bar{x}$, in which case $\partial f_\omega(\bar{x}) = \{\nabla f_\omega(\bar{x})\}$, we obtain that ϕ is differentiable at $\bar{x}$ iff conditions (i) and (ii) hold. The second assertion then follows. $\square$

Of course, if the set inside the parentheses on the right-hand side of (6.50) is closed, then there is no need to take its topological closure. This holds true in the following case.

Corollary 6.13. *Suppose that the assumptions of Theorem* 6.11 *are satisfied and for every $\zeta \in \partial\rho(\bar{Z})$ the function $f_\omega(\cdot)$ is differentiable at $\bar{x}$ for ζ-a.e. $\omega \in \Omega$. Then*

$$\partial\phi(\bar{x}) = \bigcup_{\zeta \in \partial\rho(\bar{Z})} \int_\Omega \nabla f_\omega(\bar{x})\zeta(\omega) dP(\omega). \tag{6.55}$$

Proof. In view of Theorem 6.11 we only need to show that the set on the right-hand side of (6.55) is closed. As ρ is continuous at $\bar{Z}$, the set $\partial\rho(\bar{Z})$ is weakly* compact. Also, the mapping $\zeta \mapsto \int_\Omega \nabla f_\omega(\bar{x})\zeta(\omega) dP(\omega)$, from $\mathcal{Z}^*$ to $\mathbb{R}^n$, is continuous with respect to the weak* topology of $\mathcal{Z}^*$ and the standard topology of $\mathbb{R}^n$. It follows that the image of the set $\partial\rho(\bar{Z})$ by this mapping is compact and hence is closed, i.e., the set at the right-hand side of (6.55) is closed. $\square$

6.3.2 Examples of Risk Measures

In this section we discuss several examples of risk measures which are commonly used in applications. In each of the following examples it is natural to use the space $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ for an appropriate $p \in [1, +\infty)$. Note that if a random variable Z has a pth order finite moment, then it has finite moments of any order p' smaller than p, i.e., if $1 \le p' \le p$ and $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$, then $Z \in \mathcal{L}_{p'}(\Omega, \mathcal{F}, P)$. This gives a natural embedding of $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ into $\mathcal{L}_{p'}(\Omega, \mathcal{F}, P)$ for $p' < p$. Note, however, that this embedding is not continuous. Unless stated otherwise, all expectations and probabilistic statements will be made with respect to the probability measure P.

Before proceeding to particular examples, let us consider the following construction. Let $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ and define

$$\tilde{\rho}(Z) := \mathbb{E}[Z] + \inf_{t \in \mathbb{R}} \rho(Z - t). \tag{6.56}$$

Clearly we have that for any $a \in \mathbb{R}$,

$$\tilde{\rho}(Z+a) = \mathbb{E}[Z+a] + \inf_{t\in\mathbb{R}} \rho(Z+a-t) = \mathbb{E}[Z] + a + \inf_{t\in\mathbb{R}} \rho(Z-t) = \tilde{\rho}(Z) + a.$$

That is, $\tilde{\rho}$ satisfies condition (R3) irrespective of whether ρ does. It is not difficult to see that if ρ satisfies conditions (R1) and (R2), then $\tilde{\rho}$ satisfies these conditions as well. Also, if ρ is positively homogeneous, then so is $\tilde{\rho}$. Let us calculate the conjugate of $\tilde{\rho}$. We have

$$\begin{aligned}
\tilde{\rho}^*(\zeta) &= \sup_{Z\in\mathcal{Z}} \big\{ \langle \zeta, Z\rangle - \tilde{\rho}(Z) \big\} = \sup_{Z\in\mathcal{Z}} \left\{ \langle \zeta, Z\rangle - \mathbb{E}[Z] - \inf_{t\in\mathbb{R}} \rho(Z-t) \right\} \\
&= \sup_{Z\in\mathcal{Z},\, t\in\mathbb{R}} \big\{ \langle \zeta, Z\rangle - \mathbb{E}[Z] - \rho(Z-t) \big\} \\
&= \sup_{Z\in\mathcal{Z},\, t\in\mathbb{R}} \big\{ \langle \zeta - 1, Z\rangle + t(\mathbb{E}[\zeta]-1) - \rho(Z) \big\}.
\end{aligned}$$

It follows that

$$\tilde{\rho}^*(\zeta) = \begin{cases} \rho^*(\zeta-1) & \text{if } \mathbb{E}[\zeta]=1 \\ +\infty & \text{if } \mathbb{E}[\zeta]\neq 1. \end{cases}$$

The construction below can be viewed as a homogenization of a risk measure $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$. Define

$$\check{\rho}(Z) := \inf_{\tau>0} \tau\rho(\tau^{-1}Z). \tag{6.57}$$

For any $t > 0$, by making change of variables $\tau \mapsto t\tau$, we obtain that $\check{\rho}(tZ) = t\check{\rho}(Z)$. That is, $\check{\rho}$ is positively homogeneous whether ρ is or isn't. Clearly, if ρ is positively homogeneous to start with, then $\rho = \check{\rho}$.

If ρ is convex, then so is $\check{\rho}$. Indeed, observe that if ρ is convex, then function $\varphi(\tau, Z) := \tau\rho(\tau^{-1}Z)$ is convex jointly in Z and $\tau > 0$. This can be verified directly as follows. For $t \in [0, 1]$, $\tau_1, \tau_2 > 0$, and $Z_1, Z_2 \in \mathcal{Z}$, and setting $\tau := t\tau_1 + (1-t)\tau_2$ and $Z := tZ_1 + (1-t)Z_2$, we have

$$\begin{aligned}
t[\tau_1\rho(\tau_1^{-1}Z_1)] + (1-t)[\tau_2\rho(\tau_2^{-1}Z_2)] &= \tau\left[\frac{t\tau_1}{\tau}\rho(\tau_1^{-1}Z_1) + \frac{(1-t)\tau_2}{\tau}\rho(\tau_2^{-1}Z_2)\right] \\
&\geq \tau\rho\left(\frac{t}{\tau}Z_1 + \frac{(1-t)}{\tau}Z_2\right) = \tau\rho(\tau^{-1}Z).
\end{aligned}$$

Minimizing convex function $\varphi(\tau, Z)$ over $\tau > 0$, we obtain a convex function. It is also not difficult to see that if ρ satisfies conditions (R2) and (R3), then so does $\check{\rho}$.

Let us calculate the conjugate of $\check{\rho}$. We have

$$\begin{aligned}
\check{\rho}^*(\zeta) &= \sup_{Z\in\mathcal{Z}} \big\{ \langle \zeta, Z\rangle - \check{\rho}(Z) \big\} = \sup_{Z\in\mathcal{Z},\,\tau>0} \big\{ \langle \zeta, Z\rangle - \tau\rho(\tau^{-1}Z) \big\} \\
&= \sup_{Z\in\mathcal{Z},\,\tau>0} \big\{ \tau[\langle \zeta, Z\rangle - \rho(Z)] \big\}.
\end{aligned}$$

It follows that $\check{\rho}^*$ is the indicator function of the set

$$\mathfrak{A} := \big\{ \zeta \in \mathcal{Z}^* : \langle \zeta, Z\rangle \leq \rho(Z), \quad \forall Z \in \mathcal{Z} \big\}. \tag{6.58}$$

If, moreover, $\check{\rho}$ is lower semicontinuous and then $\check{\rho}$ is equal to the conjugate of $\check{\rho}^*$, and hence $\check{\rho}$ is the support function of the above set $\mathfrak{A}$.

Example 6.14 (Utility Model). It is possible to relate the theory of convex risk measures with the utility model. Let $g : \mathbb{R} \to \overline{\mathbb{R}}$ be a proper convex nondecreasing lower semicontinuous function such that the expectation $\mathbb{E}[g(Z)]$ is well defined for all $Z \in \mathcal{Z}$. (It is allowed here for $\mathbb{E}[g(Z)]$ to take value $+\infty$ but not $-\infty$ since the corresponding risk measure is required to be proper.) We can view the function g as a *disutility* function.[45]

Proposition 6.15. *Let $g : \mathbb{R} \to \overline{\mathbb{R}}$ be a proper convex nondecreasing lower semicontinuous function. Suppose that the risk measure*

$$\rho(Z) := \mathbb{E}[g(Z)] \tag{6.59}$$

is well defined and proper. Then ρ is convex and lower semicontinuous and satisfies the monotonicity condition (R2)*, and the representation* (6.35) *holds with*

$$\rho^*(\zeta) = \mathbb{E}[g^*(\zeta)]. \tag{6.60}$$

Moreover, if $\rho(Z)$ is finite, then

$$\partial\rho(Z) = \left\{\zeta \in \mathcal{Z}^* : \zeta(\omega) \in \partial g(Z(\omega)) \ a.e.\ \omega \in \Omega\right\}. \tag{6.61}$$

Proof. Since g is lower semicontinuous and convex, we have by the Fenchel–Moreau theorem that

$$g(z) = \sup_{\alpha\in\mathbb{R}} \left\{\alpha z - g^*(\alpha)\right\},$$

where g^* is the conjugate of g. As g is proper, the conjugate function g^* is also proper. It follows that

$$\rho(Z) = \mathbb{E}\left[\sup_{\alpha\in\mathbb{R}} \left\{\alpha Z - g^*(\alpha)\right\}\right]. \tag{6.62}$$

By the interchangeability principle (Theorem 7.80) for the space $\mathfrak{M} := \mathcal{Z}^* = \mathcal{L}_q(\Omega, \mathcal{F}, P)$, which is decomposable, we obtain

$$\rho(Z) = \sup_{\zeta\in\mathcal{Z}^*} \left\{\langle\zeta, Z\rangle - \mathbb{E}[g^*(\zeta)]\right\}. \tag{6.63}$$

It follows that ρ is convex and lower semicontinuous, and representation (6.35) holds with the conjugate function given in (6.60). Moreover, since the function g is nondecreasing, it follows that ρ satisfies the monotonicity condition (R2).

Since ρ is convex proper and lower semicontinuous, and hence $\rho^{**} = \rho$, we have by Proposition 7.73 that

$$\partial\rho(Z) = \arg\max_{\zeta\in\mathfrak{A}} \left\{\mathbb{E}[\zeta Z - g^*(\zeta)]\right\}, \tag{6.64}$$

assuming that $\rho(Z)$ is finite. Together with formula (7.247) of the interchangeability principle (Theorem 7.80), this implies (6.61). $\square$

[45] We consider here minimization problems, and that is why we speak about disutility. Any disutility function g corresponds to a utility function $u : \mathbb{R} \to \mathbb{R}$ defined by $u(z) = -g(-z)$. Note that the function u is concave and nondecreasing since the function g is convex and nondecreasing.

The above risk measure ρ, defined in (6.59), does not satisfy condition (R3) unless $g(z) \equiv z$. We can consider the corresponding risk measure $\tilde{\rho}$, defined in (6.56), which in the present case can be written as

$$\tilde{\rho}(Z) = \inf_{t \in \mathbb{R}} \mathbb{E}[Z + g(Z - t)]. \tag{6.65}$$

We have that $\tilde{\rho}$ is convex since g is convex, that $\tilde{\rho}$ is monotone (i.e., condition (R2) holds) if $z + g(z)$ is monotonically nondecreasing, and that $\tilde{\rho}$ satisfies condition (R3). If, moreover, $\tilde{\rho}$ is lower semicontinuous, then the dual representation

$$\tilde{\rho}(Z) = \sup_{\substack{\zeta \in \mathcal{Z}^* \\ \mathbb{E}[\zeta]=1}} \left\{ \langle \zeta - 1, Z \rangle - \mathbb{E}[g^*(\zeta)] \right\} \tag{6.66}$$

holds. ■

Example 6.16 (Average Value-at-Risk). The risk measure ρ associated with disutility function g, defined in (6.59), is positively homogeneous only if g is positively homogeneous. Suppose now that $g(z) := \max\{az, bz\}$, where $b \geq a$. Then $g(\cdot)$ is positively homogeneous and convex. It is natural here to use the space $\mathcal{Z} := \mathcal{L}_1(\Omega, \mathcal{F}, P)$, since $\mathbb{E}[g(Z)]$ is finite for every $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$. The conjugate function of g is the indicator function $g^* = \mathbb{I}_{[a,b]}$. Therefore it follows by Proposition 6.15 that the representation (6.37) holds with

$$\mathfrak{A} = \{\zeta \in \mathcal{L}_\infty(\Omega, \mathcal{F}, P) : \zeta(\omega) \in [a, b] \text{ a.e. } \omega \in \Omega\}.$$

Note that the dual space $\mathcal{Z}^* = \mathcal{L}_\infty(\Omega, \mathcal{F}, P)$, of the space $\mathcal{Z} := \mathcal{L}_1(\Omega, \mathcal{F}, P)$, appears naturally in the corresponding representation (6.37) since, of course, the condition that "$\zeta(\omega) \in [a, b]$ for a.e. $\omega \in \Omega$" implies that ζ is essentially bounded.

Consider now the risk measure

$$\tilde{\rho}(Z) := \mathbb{E}[Z] + \inf_{t \in \mathbb{R}} \mathbb{E}\left\{\beta_1[t - Z]_+ + \beta_2[Z - t]_+\right\}, \quad Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P), \tag{6.67}$$

where $\beta_1 \in [0, 1]$ and $\beta_2 \geq 0$. This risk measure can be recognized as risk measure defined in (6.65), associated with function $g(z) := \beta_1[-z]_+ + \beta_2[z]_+$. For specified β_1 and β_2, the function $z + g(z)$ is convex and nondecreasing, and $\tilde{\rho}$ is a continuous coherent risk measure. For $\beta_1 \in (0, 1]$ and $\beta_2 > 0$, the above risk measure $\tilde{\rho}(Z)$ can be written in the form

$$\tilde{\rho}(Z) = (1 - \beta_1)\mathbb{E}[Z] + \beta_1 \mathsf{AV@R}_\alpha(Z), \tag{6.68}$$

where $\alpha := \beta_1/(\beta_1 + \beta_2)$. Note that the right-hand side of (6.67) attains its minimum at $t^* = \mathsf{V@R}_\alpha(Z)$. Therefore, the second term on the right-hand side of (6.67) is the weighted measure of deviation from the quantile $\mathsf{V@R}_\alpha(Z)$, discussed in section 6.2.3.

The respective conjugate function is the indicator function of the set $\mathfrak{A} := \text{dom}(\tilde{\rho}^*)$, and $\tilde{\rho}$ can be represented in the dual form (6.37) with

$$\mathfrak{A} = \{\zeta \in \mathcal{L}_\infty(\Omega, \mathcal{F}, P) : \zeta(\omega) \in [1 - \beta_1, 1 + \beta_2] \text{ a.e. } \omega \in \Omega,\ \mathbb{E}[\zeta] = 1\}. \tag{6.69}$$

In particular, for $\beta_1 = 1$ we have that $\tilde{\rho}(\cdot) = \mathsf{AV@R}_\alpha(\cdot)$, and hence the dual representation (6.37) of $\mathsf{AV@R}_\alpha$ holds with the set

$$\mathfrak{A} = \left\{\zeta \in \mathcal{L}_\infty(\Omega, \mathcal{F}, P) : \zeta(\omega) \in [0, \alpha^{-1}] \text{ a.e. } \omega \in \Omega,\ \mathbb{E}[\zeta] = 1\right\}. \tag{6.70}$$

Since $\mathsf{AV@R}_\alpha(\cdot)$ is convex and continuous, it is subdifferentiable and its subdifferentials can be calculated using formula (6.43). That is,

$$\partial(\mathsf{AV@R}_\alpha)(Z) = \arg\max_{\zeta \in \mathcal{Z}^*} \left\{ \langle \zeta, Z \rangle : \zeta(\omega) \in [0, \alpha^{-1}] \text{ a.e. } \omega \in \Omega,\ \mathbb{E}[\zeta] = 1 \right\}. \tag{6.71}$$

Consider the maximization problem on the right-hand side of (6.71). The Lagrangian of that problem is

$$L(\zeta, \lambda) = \langle \zeta, Z \rangle + \lambda(1 - \mathbb{E}[\zeta]) = \langle \zeta, Z - \lambda \rangle + \lambda,$$

and its (Lagrangian) dual is the problem

$$\operatorname*{Min}_{\lambda \in \mathbb{R}} \sup_{\zeta(\cdot) \in [0, \alpha^{-1}]} \left\{ \langle \zeta, Z - \lambda \rangle + \lambda \right\}. \tag{6.72}$$

We have that

$$\sup_{\zeta(\cdot) \in [0, \alpha^{-1}]} \langle \zeta, Z - \lambda \rangle = \alpha^{-1} \mathbb{E}([Z - \lambda]_+),$$

and hence the dual problem (6.72) can be written as

$$\operatorname*{Min}_{\lambda \in \mathbb{R}} \alpha^{-1} \mathbb{E}([Z - \lambda]_+) + \lambda. \tag{6.73}$$

The set of optimal solutions of problem (6.73) is the interval with the end points given by the left and right side $(1-\alpha)$-quantiles of the cdf $H_Z(z) = \mathsf{Pr}(Z \le z)$ of $Z(\omega)$. Since the set of optimal solutions of the dual problem (6.72) is a compact subset of $\mathbb{R}$, there is no duality gap between the maximization problem on the right hand side of (6.71) and its dual (6.72) (see Theorem 7.10). It follows that the set of optimal solutions of the right-hand side of (6.71), and hence the subdifferential $\partial(\mathsf{AV@R}_\alpha)(Z)$, is given by such feasible $\bar{\zeta}$ that $(\bar{\zeta}, \bar{\lambda})$ is a saddle point of the Lagrangian $L(\zeta, \lambda)$ for any $(1-\alpha)$-quantile $\bar{\lambda}$. Recall that the left-side $(1-\alpha)$-quantile of the cdf $H_Z(z)$ is called Value-at-Risk and denoted $\mathsf{V@R}_\alpha(Z)$. Suppose for the moment that the set of $(1-\alpha)$-quantiles of H_Z is a singleton, i.e., consists of one point $\mathsf{V@R}_\alpha(Z)$. Then we have

$$\partial(\mathsf{AV@R}_\alpha)(Z) = \left\{ \zeta : \mathbb{E}[\zeta] = 1,\ \begin{array}{lll} \zeta(\omega) = \alpha^{-1} & \text{if} & Z(\omega) > \mathsf{V@R}_\alpha(Z), \\ \zeta(\omega) = 0 & \text{if} & Z(\omega) < \mathsf{V@R}_\alpha(Z), \\ \zeta(\omega) \in [0, \alpha^{-1}] & \text{if} & Z(\omega) = \mathsf{V@R}_\alpha(Z). \end{array} \right. \tag{6.74}$$

If the set of $(1-\alpha)$-quantiles of H_Z is not a singleton, then the probability that $Z(\omega)$ belongs to that set is zero. Consequently, formula (6.74) still holds with the left-side quantile $\mathsf{V@R}_\alpha(Z)$ can be replaced by any $(1-\alpha)$-quantile of H_Z.

It follows that $\partial(\mathsf{AV@R}_\alpha)(Z)$ is a singleton, and hence $\mathsf{AV@R}_\alpha(\cdot)$ is Hadamard differentiable at Z, iff the following condition holds:

$$\mathsf{Pr}(Z < \mathsf{V@R}_\alpha(Z)) = 1 - \alpha \ \text{ or } \ \mathsf{Pr}(Z > \mathsf{V@R}_\alpha(Z)) = \alpha. \tag{6.75}$$

Again if the set of $(1-\alpha)$-quantiles is not a singleton, then the left-side quantile $\mathsf{V@R}_\alpha(Z)$ in the above condition (6.75) can be replaced by any $(1-\alpha)$-quantile of H_Z. Note that condition (6.75) is always satisfied if the cdf $H_Z(\cdot)$ is continuous at $\mathsf{V@R}_\alpha(Z)$, but may also hold even if $H_Z(\cdot)$ is discontinuous at $\mathsf{V@R}_\alpha(Z)$. ■

Example 6.17 (Exponential Utility Function Risk Measure). Consider utility risk measure ρ, defined in (6.59), associated with the exponential disutility function $g(z) := e^z$. That is, $\rho(Z) := \mathbb{E}[e^Z]$. A natural question is what space $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ to use here. Let us observe that unless the sigma algebra $\mathcal{F}$ has a finite number of elements, in which case $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ is finite dimensional, there exist such $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$ that $\mathbb{E}[e^Z] = +\infty$. In fact, for any $p \in [1, +\infty)$ the domain of ρ forms a dense subset of $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ and $\rho(\cdot)$ is discontinuous at every $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$ unless $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ is finite dimensional. Nevertheless, for any $p \in [1, +\infty)$ the risk measure ρ is proper and, by Proposition 6.15, is convex and lower semicontinuous. Note that if $Z : \Omega \to \mathbb{R}$ is an $\mathcal{F}$-measurable function such that $\mathbb{E}[e^Z]$ is finite, then $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$ for *any* $p \geq 1$. Therefore, by formula (6.61) of Proposition 6.15, we have that if $\mathbb{E}[e^Z]$ is finite, then $\partial\rho(Z) = \{e^Z\}$ is a singleton. It could be mentioned that although $\rho(\cdot)$ is subdifferentiable at every $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$ where it is finite and has unique subgradient e^Z, it is discontinuous and nondifferentiable at Z unless $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ is finite dimensional.

The above risk measure associated with the exponential disutility function is not positively homogeneous and does not satisfy condition (R3). Let us consider instead the risk measure

$$\rho_e(Z) := \ln \mathbb{E}[e^Z], \tag{6.76}$$

defined on $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ for some $p \in [1, +\infty)$. Since $\ln(\cdot)$ is continuous on the positive half of the real line and $\mathbb{E}[e^Z] > 0$, it follows from the above that ρ_e has the same domain as $\rho(Z) = \mathbb{E}[e^Z]$ and is lower semicontinuous and proper. It is also can be verified that ρ_e is convex. (See derivations of section 7.2.8 following (7.175).) Moreover, for any $a \in \mathbb{R}$,

$$\ln \mathbb{E}[e^{Z+a}] = \ln\left(e^a \mathbb{E}[e^Z]\right) = \ln \mathbb{E}[e^Z] + a,$$

i.e., ρ_e satisfies condition (R3).

Let us calculate the conjugate of ρ_e. We have that

$$\rho_e^*(\zeta) = \sup_{Z \in \mathcal{Z}} \left\{ \mathbb{E}[\zeta Z] - \ln \mathbb{E}[e^Z] \right\}. \tag{6.77}$$

Since ρ_e satisfies conditions (R2) and (R3), it follows that $\mathrm{dom}(\rho_e) \subset \mathfrak{P}$, where $\mathfrak{P}$ is the set of density functions (see (6.38)). By writing (first order) optimality conditions for the optimization problem on the right-hand side of (6.77), it is straightforward to verify that for $\zeta \in \mathfrak{P}$ such that $\zeta(\omega) > 0$ for a.e. $\omega \in \Omega$, a point $\bar{Z}$ is an optimal solution of that problem if $\bar{Z} = \ln \zeta + a$ for some $a \in \mathbb{R}$. Substituting this into the right-hand side of (6.77), and noting that the obtained expression does not depend on a, we obtain

$$\rho_e^*(\zeta) = \begin{cases} \mathbb{E}[\zeta \ln \zeta] & \text{if} \quad \zeta \in \mathfrak{P}, \\ +\infty & \text{if} \quad \zeta \notin \mathfrak{P}. \end{cases} \tag{6.78}$$

Note that $x \ln x$ tends to zero as $x \downarrow 0$. Therefore, we set $0 \ln 0 = 0$ in the above formula (6.78). Note also that $x \ln x$ is bounded for $x \in [0, 1]$. Therefore, $\mathrm{dom}(\rho_e^*) = \mathfrak{P}$ for any $p \in [1, +\infty)$.

Furthermore, we can apply the homogenization procedure to ρ_e (see (6.57)). That is, consider the following risk measure:

$$\check{\rho}_e(Z) := \inf_{\tau > 0} \tau \ln \mathbb{E}[e^{\tau^{-1} Z}]. \tag{6.79}$$

Risk measure $\check{\rho}_e$ satisfies conditions (R1)–(R4), i.e., it is a coherent risk measure. Its conjugate $\check{\rho}_e^*$ is the indicator function of the set (see (6.58)):

$$\mathfrak{A} := \left\{ \zeta \in \mathcal{Z}^* : \mathbb{E}[\zeta Z] \le \ln \mathbb{E}[e^Z], \quad \forall Z \in \mathcal{Z} \right\}. \tag{6.80}$$

Note that since e^z is a convex function it follows by Jensen inequality that $\mathbb{E}[Z] \le \ln \mathbb{E}[e^Z]$. Consequently, $\zeta(\cdot) = 1$ is an element of the above set $\mathfrak{A}$. ■

Example 6.18 (Mean-Variance Risk Measure). Consider

$$\rho(Z) := \mathbb{E}[Z] + c\mathbb{V}\mathrm{ar}[Z], \tag{6.81}$$

where $c \ge 0$ is a given constant. It is natural to use here the space $\mathcal{Z} := \mathcal{L}_2(\Omega, \mathcal{F}, P)$ since for any $Z \in \mathcal{L}_2(\Omega, \mathcal{F}, P)$ the expectation $\mathbb{E}[Z]$ and variance $\mathbb{V}\mathrm{ar}[Z]$ are well defined and finite. We have here that $\mathcal{Z}^* = \mathcal{Z}$ (i.e., $\mathcal{Z}$ is a Hilbert space) and for $Z \in \mathcal{Z}$ its norm is given by $\|Z\|_2 = \sqrt{\mathbb{E}[Z^2]}$. We also have that

$$\|Z\|_2^2 = \sup_{\zeta \in \mathcal{Z}} \left\{ \langle \zeta, Z \rangle - \tfrac{1}{4}\|\zeta\|_2^2 \right\}. \tag{6.82}$$

Indeed, it is not difficult to verify that the maximum on the right-hand side of (6.82) is attained at $\zeta = 2Z$.

We have that $\mathbb{V}\mathrm{ar}[Z] = \big\| Z - \mathbb{E}[Z] \big\|_2^2$, and since $\| \cdot \|_2^2$ is a convex and continuous function on the Hilbert space $\mathcal{Z}$, it follows that $\rho(\cdot)$ is convex and continuous. Also because of (6.82), we can write

$$\mathbb{V}\mathrm{ar}[Z] = \sup_{\zeta \in \mathcal{Z}} \left\{ \langle \zeta, Z - \mathbb{E}[Z] \rangle - \tfrac{1}{4}\|\zeta\|_2^2 \right\}.$$

Since

$$\langle \zeta, Z - \mathbb{E}[Z] \rangle = \langle \zeta, Z \rangle - \mathbb{E}[\zeta]\,\mathbb{E}[Z] = \langle \zeta - \mathbb{E}[\zeta], Z \rangle, \tag{6.83}$$

we can rewrite the last expression as follows:

$$\begin{aligned} \mathbb{V}\mathrm{ar}[Z] &= \sup_{\zeta \in \mathcal{Z}} \left\{ \langle \zeta - \mathbb{E}[\zeta], Z \rangle - \tfrac{1}{4}\|\zeta\|_2^2 \right\} \\ &= \sup_{\zeta \in \mathcal{Z}} \left\{ \langle \zeta - \mathbb{E}[\zeta], Z \rangle - \tfrac{1}{4}\mathbb{V}\mathrm{ar}[\zeta] - \tfrac{1}{4}\big(\mathbb{E}[\zeta]\big)^2 \right\}. \end{aligned}$$

Since $\zeta - \mathbb{E}[\zeta]$ and $\mathbb{V}\mathrm{ar}[\zeta]$ are invariant under transformations of ζ to $\zeta + a$, where $a \in \mathbb{R}$, the above maximization can be restricted to such $\zeta \in \mathcal{Z}$ that $\mathbb{E}[\zeta] = 0$. Consequently

$$\mathbb{V}\mathrm{ar}[Z] = \sup_{\substack{\zeta \in \mathcal{Z} \\ \mathbb{E}[\zeta]=0}} \left\{ \langle \zeta, Z \rangle - \tfrac{1}{4}\mathbb{V}\mathrm{ar}[\zeta] \right\}.$$

Therefore the risk measure ρ, defined in (6.81), can be expressed as

$$\rho(Z) = \mathbb{E}[Z] + c \sup_{\substack{\zeta \in \mathcal{Z} \\ \mathbb{E}[\zeta]=0}} \left\{ \langle \zeta, Z \rangle - \tfrac{1}{4}\mathbb{V}\mathrm{ar}\,[\zeta] \right\}$$

and hence for $c > 0$ (by making change of variables $\zeta' = c\zeta + 1$) as

$$\rho(Z) = \sup_{\substack{\zeta \in \mathcal{Z} \\ \mathbb{E}[\zeta]=1}} \left\{ \langle \zeta, Z \rangle - \tfrac{1}{4c} \mathbb{V}\mathrm{ar}[\zeta] \right\}. \tag{6.84}$$

It follows that for any $c > 0$ the function ρ is convex, continuous, and

$$\rho^*(\zeta) = \begin{cases} \frac{1}{4c} \mathbb{V}\mathrm{ar}[\zeta] & \text{if } \mathbb{E}[\zeta] = 1, \\ +\infty & \text{otherwise.} \end{cases} \tag{6.85}$$

The function ρ satisfies the translation equivariance condition (R3), e.g., because the domain of its conjugate contains only ζ such that $\mathbb{E}[\zeta] = 1$. However, for any $c > 0$ the function ρ is not positively homogeneous and it does not satisfy the monotonicity condition (R2), because the domain of ρ^* contains density functions which are not nonnegative.

Since $\mathbb{V}\mathrm{ar}[Z] = \langle Z, Z \rangle - (\mathbb{E}[Z])^2$, it is straightforward to verify that $\rho(\cdot)$ is (Fréchet) differentiable and

$$\nabla \rho(Z) = 2cZ - 2c\mathbb{E}[Z] + 1. \quad \blacksquare \tag{6.86}$$

Example 6.19 (Mean-Deviation Risk Measures of Order p). For $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and $\mathcal{Z}^* := \mathcal{L}_q(\Omega, \mathcal{F}, P)$, with $p \in [1, +\infty)$ and $c \geq 0$, consider

$$\rho(Z) := \mathbb{E}[Z] + c \left(\mathbb{E}\big[|Z - \mathbb{E}[Z]|^p \big] \right)^{1/p}. \tag{6.87}$$

We have that $\left(\mathbb{E}\big[|Z|^p\big]\right)^{1/p} = \|Z\|_p$, where $\|\cdot\|_p$ denotes the norm of the space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$. The function ρ is convex continuous and positively homogeneous. Also

$$\|Z\|_p = \sup_{\|\zeta\|_q \leq 1} \langle \zeta, Z \rangle, \tag{6.88}$$

and hence

$$\left(\mathbb{E}\big[|Z - \mathbb{E}[Z]|^p \big] \right)^{1/p} = \sup_{\|\zeta\|_q \leq 1} \langle \zeta, Z - \mathbb{E}[Z] \rangle = \sup_{\|\zeta\|_q \leq 1} \langle \zeta - \mathbb{E}[\zeta], Z \rangle. \tag{6.89}$$

It follows that representation (6.37) holds with the set $\mathfrak{A}$ given by

$$\mathfrak{A} = \left\{ \zeta' \in \mathcal{Z}^* : \zeta' = 1 + \zeta - \mathbb{E}[\zeta],\ \|\zeta\|_q \leq c \right\}. \tag{6.90}$$

We obtain here that ρ satisfies conditions (R1), (R3), and (R4).

The monotonicity condition (R2) is more involved. Suppose that $p = 1$. Then $q = +\infty$ and hence for any $\zeta' \in \mathfrak{A}$ and a.e. $\omega \in \Omega$ we have

$$\zeta'(\omega) = 1 + \zeta(\omega) - \mathbb{E}[\zeta] \geq 1 - |\zeta(\omega)| - \mathbb{E}[\zeta] \geq 1 - 2c.$$

It follows that if $c \in [0, 1/2]$, then $\zeta'(\omega) \geq 0$ for a.e. $\omega \in \Omega$, and hence condition (R2) follows. Conversely, take $\zeta := c(-\mathbf{1}_A + \mathbf{1}_{\Omega \setminus A})$, for some $A \in \mathcal{F}$, and $\zeta' = 1 + \zeta - \mathbb{E}[\zeta]$. We have that $\|\zeta\|_\infty = c$ and $\zeta'(\omega) = 1 - 2c + 2cP(A)$ for all $\omega \in A$ It follows that if $c > 1/2$, then $\zeta'(\omega) < 0$ for all $\omega \in A$, provided that $P(A)$ is small enough. We obtain that for $c > 1/2$ the monotonicity property (R2) does not hold if the following condition is satisfied:

$$\text{For any } \varepsilon > 0 \text{ there exists } A \in \mathcal{F} \text{ such that } \varepsilon > P(A) > 0. \tag{6.91}$$

That is, for $p = 1$ the mean-deviation measure ρ satisfies (R2) if, and provided that condition (6.91) holds, only if $c \in [0, 1/2]$. (The above condition (6.91) holds, in particular, if the measure P is nonatomic.)

Suppose now that $p > 1$. For a set $A \in \mathcal{F}$ and $\alpha > 0$ let us take $\zeta := -\alpha \mathbf{1}_A$ and $\zeta' = 1 + \zeta - \mathbb{E}[\zeta]$. Then $\|\zeta\|_q = \alpha P(A)^{1/q}$ and $\zeta'(\omega) = 1 - \alpha + \alpha P(A)$ for all $\omega \in A$. It follows that if $p > 1$, then for any $c > 0$ the mean-deviation measure ρ does not satisfy (R2) provided that condition (6.91) holds.

Since ρ is convex continuous, it is subdifferentiable. By (6.43) and because of (6.90) and (6.83) we have here that $\partial\rho(Z)$ is formed by vectors $\zeta' = 1 + \zeta - \mathbb{E}[\zeta]$ such that $\zeta \in \arg\max_{\|\zeta\|_q \le c} \langle \zeta, Z - \mathbb{E}[Z] \rangle$. That is,

$$\partial\rho(Z) = \left\{ \zeta' = 1 + c\,\zeta - c\,\mathbb{E}[\zeta] : \zeta \in \mathfrak{S}_Y \right\}, \tag{6.92}$$

where $Y(\omega) \equiv Z(\omega) - \mathbb{E}[Z]$ and $\mathfrak{S}_Y$ is the set of contact points of Y. If $p \in (1, +\infty)$, then the set $\mathfrak{S}_Y$ is a singleton, i.e., there is unique contact point ζ^*_Y, provided that $Y(\omega)$ is not zero for a.e. $\omega \in \Omega$. In that case $\rho(\cdot)$ is Hadamard differentiable at Z and

$$\nabla\rho(Z) = 1 + c\,\zeta^*_Y - c\,\mathbb{E}[\zeta^*_Y]. \tag{6.93}$$

(An explicit form of the contact point ζ^*_Y is given in (7.232).) If $Y(\omega)$ is zero for a.e. $\omega \in \Omega$, i.e., $Z(\omega)$ is constant w.p. 1, then $\mathfrak{S}_Y = \left\{ \zeta \in \mathcal{Z}^* : \|\zeta\|_q \le 1 \right\}$.

For $p = 1$ the set $\mathfrak{S}_Y$ is described in (7.233). It follows that if $p = 1$, and hence $q = +\infty$, then the subdifferential $\partial\rho(Z)$ is a singleton iff $Z(\omega) \ne \mathbb{E}[Z]$ for a.e. $\omega \in \Omega$, in which case

$$\nabla\rho(Z) = \left\{ \zeta : \begin{array}{lll} \zeta(\omega) = 1 + 2c\big(1 - \Pr(Z > \mathbb{E}[Z])\big) & \text{if} & Z(\omega) > \mathbb{E}[Z], \\ \zeta(\omega) = 1 - 2c\,\Pr(Z > \mathbb{E}[Z]) & \text{if} & Z(\omega) < \mathbb{E}[Z]. \end{array} \right. \quad \blacksquare \tag{6.94}$$

Example 6.20 (Mean-Upper-Semideviation of Order p). Let $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and for $c \ge 0$ consider[46]

$$\rho(Z) := \mathbb{E}[Z] + c\left(\mathbb{E}\left[\left[Z - \mathbb{E}[Z] \right]_+^p \right] \right)^{1/p}. \tag{6.95}$$

For any $c \ge 0$ this function satisfies conditions (R1), (R3), and (R4), and similarly to the derivations of Example 6.19 it can be shown that representation (6.37) holds with the set $\mathfrak{A}$ given by

$$\mathfrak{A} = \left\{ \zeta' \in \mathcal{Z}^* : \zeta' = 1 + \zeta - \mathbb{E}[\zeta],\ \|\zeta\|_q \le c,\ \zeta \succeq 0 \right\}. \tag{6.96}$$

Since $|\mathbb{E}[\zeta]| \le \mathbb{E}|\zeta| \le \|\zeta\|_q$ for any $\zeta \in \mathcal{L}_q(\Omega, \mathcal{F}, P)$, we have that every element of the above set $\mathfrak{A}$ is nonnegative and has its expected value equal to 1. This means that the monotonicity condition (R2) holds, if and, provided that condition (6.91) holds, only if $c \in [0, 1]$. That is, ρ is a coherent risk measure if $c \in [0, 1]$.

[46] We denote $[a]_+^p := (\max\{0, a\})^p$.

Since ρ is convex continuous, it is subdifferentiable. Its subdifferential can be calculated in a way similar to the derivations of Example 6.19. That is, $\partial\rho(Z)$ is formed by vectors $\zeta' = 1 + \zeta - \mathbb{E}[\zeta]$ such that

$$\zeta \in \arg\max \left\{\langle \zeta, Y\rangle : \|\zeta\|_q \le c,\ \zeta \succeq 0\right\}, \tag{6.97}$$

where $Y := Z - \mathbb{E}[Z]$. Suppose that $p \in (1, +\infty)$. Then the set of maximizers on the right-hand side of (6.97) is not changed if Y is replaced by Y_+, where $Y_+(\cdot) := [Y(\cdot)]_+$. Consequently, if $Z(\omega)$ is not constant for a.e. $\omega \in \Omega$, and hence $Y_+ \ne 0$, then $\partial\rho(Z)$ is a singleton and

$$\nabla\rho(Z) = 1 + c\,\zeta^*_{Y_+} - c\,\mathbb{E}[\zeta^*_{Y_+}], \tag{6.98}$$

where $\zeta^*_{Y_+}$ is the contact point of Y_+. (Note that the contact point of Y_+ is nonnegative since $Y_+ \succeq 0$.)

Suppose now that $p = 1$ and hence $q = +\infty$. Then the set on the right-hand side of (6.97) is formed by $\zeta(\cdot)$ such that $\zeta(\omega) = c$ if $Y(\omega) > 0$, $\zeta(\omega) = 0$, if $Y(\omega) < 0$, and $\zeta(\omega) \in [0, c]$ if $Y(\omega) = 0$. It follows that $\partial\rho(Z)$ is a singleton iff $Z(\omega) \ne \mathbb{E}[Z]$ for a.e. $\omega \in \Omega$, in which case

$$\nabla\rho(Z) = \begin{cases} \zeta : & \begin{array}{lll} \zeta(\omega) = 1 + c\big(1 - \Pr(Z > \mathbb{E}[Z])\big) & \text{if} & Z(\omega) > \mathbb{E}[Z], \\ \zeta(\omega) = 1 - c\Pr(Z > \mathbb{E}[Z]) & \text{if} & Z(\omega) < \mathbb{E}[Z]. \end{array}\end{cases} \tag{6.99}$$

It can be noted that by Lemma 6.1

$$\mathbb{E}\big(|Z - \mathbb{E}[Z]|\big) = 2\mathbb{E}\big([Z - \mathbb{E}[Z]]_+\big). \tag{6.100}$$

Consequently, formula (6.99) can be derived directly from (6.94). ∎

Example 6.21 (Mean-Upper-Semivariance from a Target). Let $\mathcal{Z} := \mathcal{L}_2(\Omega, \mathcal{F}, P)$ and for a weight $c \ge 0$ and a target $\tau \in \mathbb{R}$ consider

$$\rho(Z) := \mathbb{E}[Z] + c\,\mathbb{E}\left[\big[Z - \tau\big]_+^2\right]. \tag{6.101}$$

This is a convex and continuous risk measure. We can now use (6.63) with $g(z) := z + c[z - \tau]_+^2$. Since

$$g^*(\alpha) = \begin{cases} (\alpha - 1)^2/4c + \tau(\alpha - 1) & \text{if } \alpha \ge 1, \\ +\infty & \text{otherwise}, \end{cases}$$

we obtain that

$$\rho(Z) = \sup_{\zeta \in \mathcal{Z},\ \zeta(\cdot) \ge 1} \left\{\mathbb{E}[\zeta Z] - \tau\mathbb{E}[\zeta - 1] - \tfrac{1}{4c}\mathbb{E}[(\zeta - 1)^2]\right\}. \tag{6.102}$$

Consequently, representation (6.36) holds with $\mathfrak{A} = \{\zeta \in \mathcal{Z} : \zeta - 1 \succeq 0\}$ and

$$\rho^*(\zeta) = \tau\mathbb{E}[\zeta - 1] + \tfrac{1}{4c}\mathbb{E}[(\zeta - 1)^2], \quad \zeta \in \mathfrak{A}.$$

If $c > 0$, then conditions (R3) and (R4) are not satisfied by this risk measure.

Since ρ is convex continuous, it is subdifferentiable. Moreover, by using (6.61) we obtain that its subdifferentials are singletons and hence $\rho(\cdot)$ is differentiable at every $Z \in \mathcal{Z}$, and

$$\nabla\rho(Z) = \left\{\zeta : \begin{array}{lll} \zeta(\omega) = 1 + 2c(Z(\omega) - \tau) & \text{if} & Z(\omega) \geq \tau, \\ \zeta(\omega) = 1 & \text{if} & Z(\omega) < \tau. \end{array}\right. \tag{6.103}$$

The above formula can also be derived directly, and it can be shown that ρ is differentiable in the sense of Fréchet. ■

Example 6.22 (Mean-Upper-Semideviation of Order p from a Target). Let $\mathcal{Z}$ be the space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$, and for $c \geq 0$ and $\tau \in \mathbb{R}$ consider

$$\rho(Z) := \mathbb{E}[Z] + c\left(\mathbb{E}\left[[Z - \tau]_+^p\right]\right)^{1/p}. \tag{6.104}$$

For any $c \geq 0$ and τ this risk measure satisfies conditions (R1) and (R2), but not (R3) and (R4) if $c > 0$. We have

$$\begin{aligned}\left(\mathbb{E}\left[[Z - \tau]_+^p\right]\right)^{1/p} &= \sup_{\|\zeta\|_q \leq 1} \mathbb{E}\big(\zeta[Z - \tau]_+\big) = \sup_{\|\zeta\|_q \leq 1,\, \zeta(\cdot) \geq 0} \mathbb{E}\big(\zeta[Z - \tau]_+\big) \\ &= \sup_{\|\zeta\|_q \leq 1,\, \zeta(\cdot) \geq 0} \mathbb{E}\big(\zeta[Z - \tau]\big) = \sup_{\|\zeta\|_q \leq 1,\, \zeta(\cdot) \geq 0} \mathbb{E}\big[\zeta Z - \tau\zeta\big].\end{aligned}$$

We obtain that representation (6.36) holds with

$$\mathfrak{A} = \{\zeta \in \mathcal{Z}^* : \|\zeta\|_q \leq c,\ \zeta \succeq 0\}$$

and $\rho^*(\zeta) = \tau\mathbb{E}[\zeta]$ for $\zeta \in \mathfrak{A}$. ■

6.3.3 Law Invariant Risk Measures and Stochastic Orders

As in the previous sections, unless stated otherwise we assume here that $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$, $p \in [1, +\infty)$. We say that random outcomes $Z_1 \in \mathcal{Z}$ and $Z_2 \in \mathcal{Z}$ have the same distribution, with respect to the reference probability measure P, if $P(Z_1 \leq z) = P(Z_2 \leq z)$ for all $z \in \mathbb{R}$. We write this relation as $Z_1 \overset{\mathcal{D}}{\sim} Z_2$. In all examples considered in section 6.3.2, the risk measures $\rho(Z)$ discussed there were dependent only on the distribution of Z. That is, each risk measure $\rho(Z)$, considered in section 6.3.2, could be formulated in terms of the cumulative distribution function (cdf) $H_Z(t) := P(Z \leq t)$ associated with $Z \in \mathcal{Z}$. We call such risk measures law invariant (or law based, or version independent).

Definition 6.23. *A risk measure $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is* law invariant*, with respect to the reference probability measure P, if for all $Z_1, Z_2 \in \mathcal{Z}$ we have the implication*

$$\left\{Z_1 \overset{\mathcal{D}}{\sim} Z_2\right\} \Rightarrow \left\{\rho(Z_1) = \rho(Z_2)\right\}.$$

Suppose for the moment that the set $\Omega = \{\omega_1, \dots, \omega_K\}$ is finite with respective probabilities $p_1, \dots, p_K$ such that any two partial sums of p_k are different, i.e., $\sum_{k\in A} p_k =$

$\sum_{k \in B} p_k$ for $A, B \subset \{1, \dots, K\}$ iff $A = B$. Then $Z_1, Z_2 : \Omega \to \mathbb{R}$ have the same distribution only if $Z_1 = Z_2$. In that case, any risk measure, defined on the space of random variables $Z : \Omega \to \mathbb{R}$, is law invariant. Therefore, for a meaningful discussion of law invariant risk measures it is natural to consider nonatomic probability spaces.

A particular example of law invariant coherent risk measure is the Average Value-at-Risk measure $\mathsf{AV@R}_\alpha$. Clearly, a convex combination $\sum_{i=1}^m \mu_i \mathsf{AV@R}_{\alpha_i}$, with $\alpha_i \in (0, 1]$, $\mu_i \geq 0$, $\sum_{i=1}^m \mu_i = 1$, of Average Value-at-Risk measures is also a law invariant coherent risk measure. Moreover, maximum of several law invariant coherent risk measures is again a law invariant coherent risk measure. It turns out that any law invariant coherent risk measure can be constructed by the operations of taking convex combinations and maximum from the class of Average Value-at-Risk measures.

Theorem 6.24 (Kusuoka). *Suppose that the probability space $(\Omega, \mathcal{F}, P)$ is nonatomic and let $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ be a law invariant lower semicontionuous coherent risk measure. Then there exists a set $\mathfrak{M}$ of probability measures on the interval $(0, 1]$ (equipped with its Borel sigma algebra) such that*

$$\rho(Z) = \sup_{\mu \in \mathfrak{M}} \int_0^1 \mathsf{AV@R}_\alpha(Z) d\mu(\alpha), \quad \forall Z \in \mathcal{Z}. \tag{6.105}$$

In order to prove this we will need the following result.

Lemma 6.25. *Let $(\Omega, \mathcal{F}, P)$ be a nonatomic probability space and $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$. Then for $Z \in \mathcal{Z}$ and $\zeta \in \mathcal{Z}^*$ we have*

$$\sup_{Y : Y \overset{\mathcal{D}}{\sim} Z} \int_\Omega \zeta(\omega) Y(\omega) dP(\omega) = \int_0^1 H_\zeta^{-1}(t) H_Z^{-1}(t) dt, \tag{6.106}$$

where H_ζ and H_Z are the cdf's of ζ and Z, respectively.

Proof. First we prove formula (6.106) for finite set $\Omega = \{\omega_1, \dots, \omega_n\}$ with equal probabilities $P(\{\omega_i\}) = 1/n$, $i = 1, \dots, n$. For a function $Y : \Omega \to \mathbb{R}$ denote $Y_i := Y(\omega_i)$, $i = 1, \dots, n$. We have here that $Y \overset{\mathcal{D}}{\sim} Z$ iff $Y_i = Z_{\pi(i)}$ for some permutation π of the set $\{1, \dots, n\}$, and $\int_\Omega \zeta Y dP = n^{-1} \sum_{i=1}^n \zeta_i Y_i$. Moreover,[47]

$$\sum_{i=1}^n \zeta_i Y_i \leq \sum_{i=1}^n \zeta_{[i]} Y_{[i]}, \tag{6.107}$$

where $\zeta_{[1]} \leq \dots \leq \zeta_{[n]}$ are numbers $\zeta_1, \dots, \zeta_n$ arranged in the increasing order, and $Y_{[1]} \leq \cdots Y_{[n]}$ are numbers $Y_1, \dots, Y_n$ arranged in the increasing order. It follows that

$$\sup_{Y : Y \overset{\mathcal{D}}{\sim} Z} \int_\Omega \zeta(\omega) Y(\omega) dP(\omega) = n^{-1} \sum_{i=1}^n \zeta_{[i]} Z_{[i]}. \tag{6.108}$$

[47] Inequality (6.107) is called the Hardy–Littlewood–Polya inequality (compare with the proof of Theorem 4.50).

It remains to note that in the considered case the right-hand side of (6.108) coincides with the right-hand side of (6.106).

Now if the space $(\Omega, \mathcal{F}, P)$ is nonatomic, we can partition Ω into n disjoint subsets, each of the same P-measure $1/n$, and it suffices to verify formula (6.106) for functions which are piecewise constant on such partitions. This reduces the problem to the case considered above. □

Proof of Theorem 6.24. By the dual representation (6.37) of Theorem 6.4, we have that for $Z \in \mathcal{Z}$,

$$\rho(Z) = \sup_{\zeta \in \mathfrak{A}} \int_{\Omega} \zeta(\omega) Z(\omega) dP(\omega), \tag{6.109}$$

where $\mathfrak{A}$ is a set of probability density functions in $\mathcal{Z}^*$. Since ρ is law invariant, we have that

$$\rho(Z) = \sup_{Y \in \mathfrak{D}(Z)} \rho(Y),$$

where $\mathfrak{D}(Z) := \{Y \in \mathcal{Z} : Y \overset{\mathcal{D}}{\sim} Z\}$. Consequently,

$$\rho(Z) = \sup_{Y \in \mathfrak{D}(Z)} \left[\sup_{\zeta \in \mathfrak{A}} \int_{\Omega} \zeta(\omega) Y(\omega) dP(\omega) \right] = \sup_{\zeta \in \mathfrak{A}} \left[\sup_{Y \in \mathfrak{D}(Z)} \int_0^1 \zeta(\omega) Y(\omega) dP(\omega) \right]. \tag{6.110}$$

Moreover, by Lemma 6.25 we have

$$\sup_{Y \in \mathfrak{D}(Z)} \int_{\Omega} \zeta(\omega) Y(\omega) dP(\omega) = \int_0^1 H_\zeta^{-1}(t) H_Z^{-1}(t) dt, \tag{6.111}$$

where H_ζ and H_Z are the cdf's of $\zeta(\omega)$ and $Z(\omega)$, respectively.

Recalling that $H_Z^{-1}(t) = \mathsf{V@R}_{1-t}(Z)$, we can write (6.111) in the form

$$\sup_{Y \in \mathfrak{D}(Z)} \int_{\Omega} \zeta(\omega) Y(\omega) dP(\omega) = \int_0^1 H_\zeta^{-1}(t) \mathsf{V@R}_{1-t}(Z) dt, \tag{6.112}$$

which together with (6.110) imply that

$$\rho(Z) = \sup_{\zeta \in \mathfrak{A}} \int_0^1 H_\zeta^{-1}(t) \mathsf{V@R}_{1-t}(Z) dt. \tag{6.113}$$

For $\zeta \in \mathfrak{A}$, the function $H_\zeta^{-1}(t)$ is monotonically nondecreasing on [0,1] and can be represented in the form

$$H_\zeta^{-1}(t) = \int_{1-t}^1 \alpha^{-1} d\mu(\alpha) \tag{6.114}$$

for some measure μ on [0,1]. Moreover, for $\zeta \in \mathfrak{A}$ we have that $\int \zeta dP = 1$, and hence $\int_0^1 H_\zeta^{-1}(t) dt = \int \zeta dP = 1$, and therefore

$$1 = \int_0^1 \int_{1-t}^1 \alpha^{-1} d\mu(\alpha) dt = \int_0^1 \int_{1-\alpha}^1 \alpha^{-1} dt d\mu(\alpha) = \int_0^1 d\mu(\alpha).$$

Consequently, μ is a probability measure on [0,1]. Also (see Theorem 6.2) we have

$$\mathsf{AV@R}_\alpha(Z) = \frac{1}{\alpha}\int_{1-\alpha}^{1} \mathsf{V@R}_{1-t}(Z)dt,$$

and hence

$$\begin{array}{rcl} \int_0^1 \mathsf{AV@R}_\alpha(Z)d\mu(\alpha) & = & \int_0^1 \int_{1-\alpha}^1 \alpha^{-1}\mathsf{V@R}_{1-t}(Z)dtd\mu(\alpha) \\ & = & \int_0^1 \mathsf{V@R}_{1-t}(Z)\left(\int_{1-t}^1 \alpha^{-1}d\mu(\alpha)\right)dt \\ & = & \int_0^1 \mathsf{V@R}_{1-t}(Z)\, H_\zeta^{-1}(t)dt. \end{array}$$

By (6.113) this completes the proof, with the correspondence between $\zeta \in \mathfrak{A}$ and $\mu \in \mathfrak{M}$ given by (6.114). □

Example 6.26. Consider $\rho := \mathsf{AV@R}_\gamma$ risk measure for some $\gamma \in (0, 1)$. Assume that the corresponding probability space is $\Omega = [0, 1]$ equipped with its Borel sigma algebra and uniform probability measure P. We have here (see (6.70))

$$\mathfrak{A} = \left\{\zeta : 0 \le \zeta(\omega) \le \gamma^{-1},\ \omega \in [0,1],\ \int_0^1 \zeta(\omega)d\omega = 1\right\}.$$

Consequently, the family of cumulative distribution functions H_ζ^{-1}, $\zeta \in \mathfrak{A}$, is formed by left-side continuous monotonically nondecreasing on [0,1] functions with $\int_0^1 H_\zeta^{-1}(t)dt = 1$ and range values $0 \le H_\zeta^{-1}(t) \le \gamma^{-1}$, $t \in [0, 1]$. Since $\mathsf{V@R}_{1-t}(Z)$ is monotonically nondecreasing in t function, it follows that the maximum in the right-hand side of (6.113) is attained at $\zeta \in \mathfrak{A}$ such that $H_\zeta^{-1}(t) = 0$ for $t \in [0, 1-\gamma]$, and $H_\zeta^{-1}(t) = \gamma^{-1}$ for $t \in (1-\gamma, 1]$. The corresponding measure μ, defined by (6.114), is given by function $\mu(\alpha) = 1$ for $\alpha \in [0, \gamma]$ and $\mu(\alpha) = 1$ for $\alpha \in (\gamma, 1]$, i.e., μ is the measure of mass 1 at the point γ. By the above proof of Theorem 6.24, this μ is the maximizer of the right-hand side of (6.105). It follows that the representation (6.105) recovers the measure $\mathsf{AV@R}_\gamma$, as it should be. ∎

For law invariant risk measures, it makes sense to discuss their monotonicity properties with respect to various stochastic orders defined for (real valued) random variables. Many stochastic orders can be characterized by a class $\mathcal{U}$ of functions $u : \mathbb{R} \to \mathbb{R}$ as follows. For (real valued) random variables Z_1 and Z_2 it is said that Z_2 dominates Z_1, denoted $Z_2 \succeq_u Z_1$, if $\mathbb{E}[u(Z_2)] \ge \mathbb{E}[u(Z_1)]$ for all $u \in \mathcal{U}$ for which the corresponding expectations do exist. This stochastic order is called the *integral stochastic order* with *generator* $\mathcal{U}$. In particular, the *usual stochastic order*, written $Z_2 \succeq_{(1)} Z_1$, corresponds to the generator $\mathcal{U}$ formed by all nondecreasing functions $u : \mathbb{R} \to \mathbb{R}$. Equivalently, $Z_2 \succeq_{(1)} Z_1$ iff $H_{Z_2}(t) \le H_{Z_1}(t)$ for all $t \in \mathbb{R}$. The relation $\succeq_{(1)}$ is also frequently called the *first order stochastic dominance* (see Definition 4.3). We say that the integral stochastic order is *increasing* if all functions in the set $\mathcal{U}$ are nondecreasing. The usual stochastic order is an example of increasing integral stochastic order.

Definition 6.27. *A law invariant risk measure $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is* consistent *(monotone) with the integral stochastic order $\succeq_u$ if for all $Z_1, Z_2 \in \mathcal{Z}$ we have the implication*

$$\left\{Z_2 \succeq_u Z_1\right\} \Rightarrow \left\{\rho(Z_2) \ge \rho(Z_1)\right\}.$$

For an increasing integral stochastic order we have that if $Z_2(\omega) \geq Z_1(\omega)$ for a.e. $\omega \in \Omega$, then $u(Z_2(\omega)) \geq u(Z_1(\omega))$ for any $u \in \mathcal{U}$ and a.e. $\omega \in \Omega$, and hence $\mathbb{E}[u(Z_2)] \geq \mathbb{E}[u(Z_1)]$. That is, if $Z_2 \succeq Z_1$ in the almost sure sense, then $Z_2 \succeq_{\mathcal{U}} Z_1$. It follows that if ρ is law invariant and consistent with respect to an increasing integral stochastic order, then it satisfies the monotonicity condition (R2). In other words, if ρ does not satisfy condition (R2), then it cannot be consistent with any increasing integral stochastic order. In particular, for $c > 1$ the mean-semideviation risk measure, defined in Example 6.20, is not consistent with any increasing integral stochastic order, provided that condition (6.91) holds.

A general way of proving consistency of law invariant risk measures with stochastic orders can be obtained via the following construction. For a given pair of random variables Z_1 and Z_2 in $\mathcal{Z}$, consider another pair of random variables, $\hat{Z}_1$ and $\hat{Z}_2$, which have distributions identical to the original pair, i.e., $\hat{Z}_1 \overset{\mathcal{D}}{\sim} Z_1$ and $\hat{Z}_2 \overset{\mathcal{D}}{\sim} Z_2$. The construction is such that the postulated consistency result becomes evident. For this method to be applicable, it is convenient to assume that the probability space $(\Omega, \mathcal{F}, P)$ is nonatomic. Then there exists a measurable function $U : \Omega \to \mathbb{R}$ (uniform random variable) such that $P(U \leq t) = t$ for all $t \in [0, 1]$.

Theorem 6.28. *Suppose that the probability space $(\Omega, \mathcal{F}, P)$ is nonatomic. Then the following holds: if a risk measure $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is law invariant, then it is consistent with the usual stochastic order iff it satisfies the monotonicity condition* (R2).

Proof. By the discussion preceding the theorem, it is sufficient to prove that (R2) implies consistency with the usual stochastic order.

For a uniform random variable $U(\omega)$ consider the random variables $\hat{Z}_1 := H_{Z_1}^{-1}(U)$ and $\hat{Z}_2 := H_{Z_2}^{-1}(U)$. We obtain that if $Z_2 \succeq_{(1)} Z_1$, then $\hat{Z}_2(\omega) \geq \hat{Z}_1(\omega)$ for all $\omega \in \Omega$, and hence by virtue of (R2), $\rho(\hat{Z}_2) \geq \rho(\hat{Z}_1)$. By construction, $\hat{Z}_1 \overset{\mathcal{D}}{\sim} Z_1$ and $\hat{Z}_2 \overset{\mathcal{D}}{\sim} Z_2$. Since the risk measure is law invariant, we conclude that $\rho(Z_2) \geq \rho(Z_1)$. Consequently, the risk measure ρ is consistent with the usual stochastic order. $\square$

It is said that Z_1 is smaller than Z_2 in the *increasing convex order*, written $Z_1 \preceq_{\text{icx}} Z_2$, if $\mathbb{E}[u(Z_1)] \leq \mathbb{E}[u(Z_2)]$ for all increasing convex functions $u : \mathbb{R} \to \mathbb{R}$ such that the expectations exist. Clearly this is an integral stochastic order with the corresponding generator given by the set of increasing convex functions. It is equivalent to the *second order stochastic dominance* relation for the negative variables: $-Z_1 \succeq_{(2)} -Z_2$. (Recall that we are dealing here with minimization rather than maximization problems.) Indeed, applying Definition 4.4 to $-Z_1$ and $-Z_2$ for $k = 2$ and using identity (4.7) we see that

$$\mathbb{E}\big\{[Z_1 - \eta]_+\big\} \leq \mathbb{E}\big\{[Z_2 - \eta]_+\big\}, \quad \forall \eta \in \mathbb{R}. \tag{6.115}$$

Since any convex nondecreasing function $u(z)$ can be arbitrarily close approximated by a positive combination of functions $u_k(z) = \beta_k + [z - \eta_k]_+$, inequality (6.115) implies that $\mathbb{E}[u(Z_1)] \leq \mathbb{E}[u(Z_2)]$, as claimed (compare with the statement (4.8)).

Theorem 6.29. *Suppose that the probability space $(\Omega, \mathcal{F}, P)$ is nonatomic. Then any law invariant lower semicontinuous coherent risk measure $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is consistent with the increasing convex order.*

Proof. By using definition (6.22) of $\mathsf{AV@R}_\alpha$ and the property that $Z_1 \preceq_{\rm icx} Z_2$ iff condition (6.115) holds, it is straightforward to verify that $\mathsf{AV@R}_\alpha$ is consistent with the increasing convex order. Now by using the representation (6.105) of Theorem 6.24 and noting that the operations of taking convex combinations and maximum preserve consistency with the increasing convex order, we can complete the proof. □

Remark 20. For *convex* risk measures (without the positive homogeneity property), Theorem 6.29 in the space $\mathcal{L}_1(\Omega, \mathcal{F}, P)$ can be derived from Theorem 4.52, which for the increasing convex order can be written as follows:

$$\{Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P) : Z \preceq_{\rm icx} Y\} = \operatorname{cl}\operatorname{conv}\{Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P) : Z \preceq_{(1)} Y\}. \tag{6.116}$$

If Z is an element of the set in the left-hand side of (6.116), then there exists a sequence of random variables $Z^k \to Z$, which are convex combinations of some elements of the set in the right-hand side of (6.116), that is,

$$Z^k = \sum_{j=1}^{N_k} \alpha_j^k Z_j^k, \qquad \sum_{j=1}^{N_k} \alpha_j^k = 1, \qquad \alpha_j^k \ge 0, \qquad Z_j^k \preceq_{(1)} Y.$$

By convexity of ρ and by Theorem 6.28, we obtain

$$\rho(Z^k) \le \sum_{j=1}^{N_k} \alpha_j^k \rho(Z_j^k) \le \sum_{j=1}^{N_k} \alpha_j^k \rho(Y) = \rho(Y).$$

Passing to the limit with $k \to \infty$ and using lower semicontinuity of ρ, we obtain $\rho(Z) \le \rho(Y)$, as required.

If $p > 1$ the domain of ρ can be extended to $\mathcal{L}_1(\Omega, \mathcal{F}, P)$, while preserving its lower semicontinuity (cf. Filipović and Svindland [66]).

Remark 21. For some measures of risk, in particular, for the mean-semideviation measures, defined in Example 6.20, and for the Average Value-at-Risk, defined in Example 6.16, consistency with the increasing convex order can be proved *without* the assumption that the probability space $(\Omega, \mathcal{F}, P)$ is nonatomic by using the following construction. Let $(\Omega, \mathcal{F}, P)$ be a nonatomic probability space; for example, we can take Ω as the interval $[0, 1]$ equipped with its Borel sigma algebra and uniform probability measure P. Then for any finite set of probabilities $p_k > 0$, $k = 1, \dots, K$, $\sum_{i=1}^K p_k = 1$, we can construct a partition of the set $\Omega = \cup_{k=1}^K A_k$ such that $P(A_k) = p_k$, $k = 1, \dots, K$. Consider the linear subspace of the respective space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ formed by piecewise constant on the sets A_k functions $Z : \Omega \to \mathbb{R}$. We can identify this subspace with the space of random variables defined on a finite probability space of cardinality K with the respective probabilities p_k, $k = 1, \dots, K$. By the above theorem, the mean-upper-semideviation risk measure (of order p) defined on $(\Omega, \mathcal{F}, P)$ is consistent with the increasing convex order. This property is preserved by restricting it to the constructed subspace. This shows that the mean-upper-semideviation risk measures are consistent with the increasing convex order on any finite probability space. This can be extended to the general probability spaces by continuity arguments.

Corollary 6.30. *Suppose that the probability space $(\Omega, \mathcal{F}, P)$ is nonatomic. Let $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ be a law invariant lower semicontinuous coherent risk measure and $\mathcal{G}$ be a sigma subalgebra of the sigma algebra $\mathcal{F}$. Then*

$$\rho\left(\mathbb{E}\left[Z|\mathcal{G}\right]\right) \le \rho(Z), \quad \forall Z \in \mathcal{Z}, \tag{6.117}$$

and

$$\mathbb{E}\left[Z\right] \le \rho(Z), \quad \forall Z \in \mathcal{Z}. \tag{6.118}$$

Proof. Consider $Z \in \mathcal{Z}$ and $Z' := \mathbb{E}[Z|\mathcal{G}]$. For every convex function $u : \mathbb{R} \to \mathbb{R}$ we have

$$\mathbb{E}[u(Z')] = \mathbb{E}\left[u\left(\mathbb{E}[Z|\mathcal{G}]\right)\right] \le \mathbb{E}[\mathbb{E}(u(Z)|\mathcal{G})] = \mathbb{E}\left[u(Z)\right],$$

where the inequality is implied by Jensen's inequality. This shows that $Z' \preceq_{\text{icx}} Z$, and hence (6.117) follows by Theorem 6.29.

In particular, for $\mathcal{G} := \{\Omega, \emptyset\}$, it follows by (6.117) that $\rho(Z) \ge \rho\left(\mathbb{E}\left[Z\right]\right)$, and since $\rho\left(\mathbb{E}\left[Z\right]\right) = \mathbb{E}[Z]$ this completes the proof. $\square$

An intuitive interpretation of property (6.117) is that if we reduce variability of a random variable Z by employing conditional averaging $Z' = \mathbb{E}[Z|\mathcal{G}]$, then the risk measure $\rho(Z')$ becomes smaller, while $\mathbb{E}[Z'] = \mathbb{E}[Z]$.

6.3.4 Relation to Ambiguous Chance Constraints

Owing to the dual representation (6.36), measures of risk are related to robust and ambiguous models. Consider a chance constraint of the form

$$P\{C(x, \omega) \le 0\} \ge 1 - \alpha. \tag{6.119}$$

Here P is a probability measure on a measurable space $(\Omega, \mathcal{F})$ and $C : \mathbb{R}^n \times \Omega \to \mathbb{R}$ is a random function. It is assumed in this formulation of chance constraint that the probability measure (distribution), with respect to which the corresponding probabilities are calculated, is known. Suppose now that the underlying probability distribution is not known exactly but rather is assumed to belong to a specified family of probability distributions. Problems involving such constrains are called *ambiguous chance constrained* problems. For a specified uncertainty set $\mathfrak{A}$ of probability measures on $(\Omega, \mathcal{F})$, the corresponding ambiguous chance constraint defines a feasible set $X \subset \mathbb{R}^n$, which can be written as

$$X := \left\{x : \mu\{C(x, \omega) \le 0\} \ge 1 - \alpha, \quad \forall \mu \in \mathfrak{A}\right\}. \tag{6.120}$$

The set X can be written in the following equivalent form:

$$X = \left\{x \in \mathbb{R}^n : \sup_{\mu \in \mathfrak{A}} \mathbb{E}_\mu\left[\mathbf{1}_{A_x}\right] \le \alpha\right\}, \tag{6.121}$$

where $A_x := \{\omega \in \Omega : C(x, \omega) > 0\}$. Recall that by the duality representation (6.37), with the set $\mathfrak{A}$ is associated a coherent risk measure ρ, and hence (6.121) can be written as

$$X = \left\{x \in \mathbb{R}^n : \rho\left(\mathbf{1}_{A_x}\right) \le \alpha\right\}. \tag{6.122}$$

We discuss now constraints of the form (6.122) where the respective risk measure is defined in a direct way. As before, we use spaces $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$, where P is viewed as a reference probability measure.

It is not difficult to see that if ρ is a law invariant risk measure, then for $A \in \mathcal{F}$ the quantity $\rho(\mathbf{1}_A)$ depends only on $P(A)$. Indeed, if $Z := \mathbf{1}_A$ for some $A \in \mathcal{F}$, then its cdf $H_Z(z) := P(Z \le z)$ is

$$H_Z(z) = \begin{cases} 0 & \text{if} \quad z < 0, \\ 1 - P(A) & \text{if} \quad 0 \le z < 1, \\ 1 & \text{if} \quad 1 \le z, \end{cases}$$

which clearly depends only on $P(A)$.

- With every law invariant real valued risk measure $\rho : \mathcal{Z} \to \mathbb{R}$ we associate function φ_ρ defined as $\varphi_\rho(t) := \rho(\mathbf{1}_A)$, where $A \in \mathcal{F}$ is any event such that $P(A) = t$, and $t \in T := \{P(A) : A \in \mathcal{F}\}$.

The function φ_ρ is well defined because for law invariant risk measure ρ the quantity $\rho(\mathbf{1}_A)$ depends only on the probability $P(A)$ and hence $\rho(\mathbf{1}_A)$ is the same for any $A \in \mathcal{F}$ such that $P(A) = t$ for a given $t \in T$. Clearly T is a subset of the interval $[0, 1]$, and $0 \in T$ (since $\emptyset \in \mathcal{F}$) and $1 \in T$ (since $\Omega \in \mathcal{F}$). If P is a nonatomic measure, then for any $A \in \mathcal{F}$ the set $\{P(B) : B \subset A,\ B \in \mathcal{F}\}$ coincides with the interval $[0, P(A)]$. In particular, if P is nonatomic, then the set $T = \{P(A) : A \in \mathcal{F}\}$, on which φ_ρ is defined, coincides with the interval $[0, 1]$.

Proposition 6.31. *Let $\rho : \mathcal{Z} \to \mathbb{R}$ be a (real valued) law invariant coherent risk measure. Suppose that the reference probability measure P is nonatomic. Then $\varphi_\rho(\cdot)$ is a continuous nondecreasing function defined on the interval* $[0, 1]$ *such that $\varphi_\rho(0) = 0$ and $\varphi_\rho(1) = 1$, and $\varphi_\rho(t) \ge t$ for all $t \in [0, 1]$.*

Proof. Since the coherent risk measure ρ is real valued, it is continuous. Because ρ is continuous and positively homogeneous, $\rho(0) = 0$ and hence $\varphi_\rho(0) = 0$. Also by (R3), we have that $\rho(\mathbf{1}_\Omega) = 1$ and hence $\varphi_\rho(1) = 1$. By Corollary 6.30 we have that $\rho(\mathbf{1}_A) \ge P(A)$ for any $A \in \mathcal{F}$ and hence $\varphi_\rho(t) \ge t$ for all $t \in [0, 1]$.

Let $t_k \in [0, 1]$ be a monotonically increasing sequence tending to t^*. Since P is a nonatomic, there exists a sequence $A_1 \subset A_2 \subset \cdots$ of $\mathcal{F}$-measurable sets such that $P(A_k) = t_k$ for all $k \in \mathbb{N}$. It follows that the set $A := \cup_{k=1}^\infty A_k$ is $\mathcal{F}$-measurable and $P(A) = t^*$. Since $\mathbf{1}_{A_k}$ converges (in the norm topology of $\mathcal{Z}$) to $\mathbf{1}_A$, it follows by continuity of ρ that $\rho(\mathbf{1}_{A_k})$ tends to $\rho(\mathbf{1}_A)$, and hence $\varphi_\rho(t_k)$ tends to $\varphi_\rho(t^*)$. In a similar way we have that $\varphi_\rho(t_k) \to \varphi_\rho(t^*)$ for a monotonically decreasing sequence t_k tending to t^*. This shows that φ_ρ is continuous.

For any $0 \le t_1 < t_2 \le 1$ there exist sets $A, B \in \mathcal{F}$ such that $B \subset A$ and $P(B) = t_1$, $P(A) = t_2$. Since $\mathbf{1}_A \succeq \mathbf{1}_B$, it follows by monotonicity of ρ that $\rho(\mathbf{1}_A) \ge \rho(\mathbf{1}_B)$. This implies that $\varphi_\rho(t_2) \ge \varphi_\rho(t_1)$, i.e., φ_ρ is nondecreasing. $\square$

Now consider again the set X of the form (6.120). Assuming conditions of Proposition 6.31, we obtain that this set X can be written in the following equivalent form:

$$X = \big\{x : P\{C(x, \omega) \le 0\} \ge 1 - \alpha^*\big\}, \tag{6.123}$$

where $\alpha^* := \varphi_\rho^{-1}(\alpha)$. That is, X can be defined by a chance constraint with respect to the reference distribution P and with the respective significance level α^*. Since $\varphi_\rho(t) \ge t$, for any $t \in [0, 1]$, it follows that $\alpha^* \le \alpha$. Let us consider some examples.

Consider Average Value-at-Risk measure $\rho := \mathsf{AV@R}_\gamma$, $\gamma \in (0, 1]$. By direct calculations it is straightforward to verify that for any $A \in \mathcal{F}$

$$\mathsf{AV@R}_\gamma(\mathbf{1}_A) = \begin{cases} \gamma^{-1} P(A) & \text{if} \quad P(A) \le \gamma, \\ 1 & \text{if} \quad P(A) > \gamma. \end{cases}$$

Consequently the corresponding function $\varphi_\rho(t) = \gamma^{-1} t$ for $t \in [0, \gamma]$, and $\varphi_\rho(t) = 1$ for $t \in [\gamma, 1]$. Now let ρ be a convex combination of Average Value-at-Risk measures, i.e., $\rho := \sum_{i=1}^m \lambda_i \rho_i$, with $\rho_i := \mathsf{AV@R}_{\gamma_i}$ and positive weights λ_i summing up to one. By the definition of the function φ_ρ we have then that $\varphi_\rho = \sum_{i=1}^m \lambda_i \varphi_{\rho_i}$. It follows that $\varphi_\rho : [0, 1] \to [0, 1]$ is a piecewise linear nondecreasing concave function with $\varphi_\rho(0) = 0$ and $\varphi_\rho(1) = 1$. More generally, let λ be a probability measure on $(0, 1]$ and $\rho := \int_0^1 \mathsf{AV@R}_\gamma d\lambda(\gamma)$. In that case, the corresponding function φ_ρ becomes a nondecreasing concave function with $\varphi_\rho(0) = 0$ and $\varphi_\rho(1) = 1$. We also can consider measures ρ given by the maximum of such integral functions over some set $\mathfrak{M}$ of probability measures on $(0, 1]$. In that case the respective function φ_ρ becomes the maximum of the corresponding nondecreasing concave functions. By Theorem 6.24 this actually gives the most general form of the function φ_ρ.

For instance, let $\mathcal{Z} := \mathcal{L}_1(\Omega, \mathcal{F}, P)$ and $\rho(Z) := (1 - \beta)\mathbb{E}[Z] + \beta \mathsf{AV@R}_\gamma(Z)$, where $\beta, \gamma \in (0, 1)$ and the expectations are taken with respect to the reference distribution P. This risk measure was discussed in example 6.16. Then

$$\varphi_\rho(t) = \begin{cases} (1 - \beta + \gamma^{-1}\beta)t & \text{if } t \in [0, \gamma], \\ \beta + (1 - \beta)t & \text{if } t \in (\gamma, 1]. \end{cases} \tag{6.124}$$

It follows that for this risk measure and for $\alpha \le \beta + (1 - \beta)\gamma$,

$$\alpha^* = \frac{\alpha}{1 + \beta(\gamma^{-1} - 1)}. \tag{6.125}$$

In particular, for $\beta = 1$, i.e., for $\rho = \mathsf{AV@R}_\gamma$, we have that $\alpha^* = \gamma\alpha$.

As another example consider the mean-upper-semideviation risk measure of order p. That is, $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and

$$\rho(Z) := \mathbb{E}[Z] + c\left(\mathbb{E}\left[\left[Z - \mathbb{E}[Z]\right]_+^p\right]\right)^{1/p}$$

(see Example 6.20). We have here that $\rho(\mathbf{1}_A) = P(A) + c[P(A)(1 - P(A))^p]^{1/p}$, and hence

$$\varphi_\rho(t) = t + c\, t^{1/p}(1 - t), \quad t \in [0, 1]. \tag{6.126}$$

In particular, for $p = 1$ we have that $\varphi_\rho(t) = (1 + c)t - ct^2$, and hence

$$\alpha^* = \frac{1 + c - \sqrt{(1 + c)^2 - 4\alpha c}}{2c}. \tag{6.127}$$

Note that for $c > 1$ the above function $\varphi_\rho(\cdot)$ is not monotonically nondecreasing on the interval $[0, 1]$. This should be not surprising since for $c > 1$ and nonatomic P, the corresponding mean-upper-semideviation risk measure is not monotone.

6.4 Optimization of Risk Measures

As before, we use spaces $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and $\mathcal{Z}^* = \mathcal{L}_q(\Omega, \mathcal{F}, P)$. Consider the composite function $\phi(\cdot) := \rho(F(\cdot))$, also denoted $\phi = \rho \circ F$, associated with a mapping $F : \mathbb{R}^n \to \mathcal{Z}$ and a risk measure $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$. We already studied properties of such composite functions in section 6.3.1. Again we write $f(x, \omega)$ or $f_\omega(x)$ for $[F(x)](\omega)$ and view $f(x, \omega)$ as a random function defined on the measurable space $(\Omega, \mathcal{F})$. Note that $F(x)$ is an element of space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ and hence $f(x, \cdot)$ is $\mathcal{F}$-measurable and finite valued. If, moreover, $f(\cdot, \omega)$ is continuous for a.e. $\omega \in \Omega$, then $f(x, \omega)$ is a Carathéodory function, and hence is random lower semicontinuous.

In this section we discuss optimization problems of the form

$$\operatorname*{Min}_{x \in X} \left\{ \phi(x) := \rho(F(x)) \right\}. \tag{6.128}$$

Unless stated otherwise, we assume that the feasible set X is a nonempty convex closed subset of $\mathbb{R}^n$. Of course, if we use $\rho(\cdot) := \mathbb{E}[\cdot]$, then problem (6.128) becomes a standard stochastic problem of optimizing (minimizing) the expected value of the random function $f(x, \omega)$. In that case we can view the corresponding optimization problem as *risk neutral*. However, a particular realization of $f(x, \omega)$ could be quite different from its expectation $\mathbb{E}[f(x, \omega)]$. This motivates an introduction, in the corresponding optimization procedure, of some type of risk control. In the analysis of portfolio selection (see section 1.4), we discussed an approach of using variance as a measure of risk. There is, however, a problem with such approach since the corresponding mean-variance risk measure is not monotone (see Example 6.18). We shall discuss this later.

Unless stated otherwise we assume that the risk measure ρ is proper and lower semicontinuous and satisfies conditions (R1)–(R2). By Theorem 6.4 we can use representation (6.36) to write problem (6.128) in the form

$$\operatorname*{Min}_{x \in X} \sup_{\zeta \in \mathfrak{A}} \Phi(x, \zeta), \tag{6.129}$$

where $\mathfrak{A} := \operatorname{dom}(\rho^*)$ and the function $\Phi : \mathbb{R}^n \times \mathcal{Z}^* \to \overline{\mathbb{R}}$ is defined by

$$\Phi(x, \zeta) := \int_\Omega f(x, \omega)\zeta(\omega)dP(\omega) - \rho^*(\zeta). \tag{6.130}$$

If, moreover, ρ is positively homogeneous, then ρ^* is the indicator function of the set $\mathfrak{A}$ and hence $\rho^*(\cdot)$ is identically zero on $\mathfrak{A}$. That is, if ρ is a proper lower semicontinuous coherent risk measure, then problem (6.128) can be written as the minimax problem

$$\operatorname*{Min}_{x \in X} \sup_{\zeta \in \mathfrak{A}} \mathbb{E}_\zeta[f(x, \omega)], \tag{6.131}$$

where

$$\mathbb{E}_\zeta[f(x, \omega)] := \int_\Omega f(x, \omega)\zeta(\omega)dP(\omega)$$

denotes the expectation with respect to ζdP. Note that, by the definition, $F(x) \in \mathcal{Z}$ and $\zeta \in \mathcal{Z}^*$, and hence

$$\mathbb{E}_\zeta[f(x, \omega)] = \langle F(x), \zeta \rangle$$

is finite valued.

Suppose that the mapping $F : \mathbb{R}^n \to \mathcal{Z}$ is convex, i.e., for a.e. $\omega \in \Omega$ the function $f(\cdot, \omega)$ is convex. This implies that for every $\zeta \succeq 0$ the function $\Phi(\cdot, \zeta)$ is convex and if, moreover, $\zeta \in \mathfrak{A}$, then $\Phi(\cdot, \zeta)$ is real valued and hence continuous. We also have that $\langle F(x), \zeta \rangle$ is linear and $\rho^*(\zeta)$ is convex in $\zeta \in \mathcal{Z}^*$, and hence for every $x \in X$ the function $\Phi(x, \cdot)$ is concave. Therefore, under various regularity conditions, there is no duality gap between problem (6.128) and its dual

$$\operatorname*{Max}_{\zeta \in \mathfrak{A}} \inf_{x \in X} \left\{ \int_\Omega f(x, \omega)\zeta(\omega) dP(\omega) - \rho^*(\zeta) \right\}, \tag{6.132}$$

which is obtained by interchanging the min and max operators in (6.129). (Recall that the set X is assumed to be nonempty closed and convex.) In particular, if there exists a saddle point $(\bar{x}, \bar{\zeta}) \in X \times \mathfrak{A}$ of the minimax problem (6.129), then there is no duality gap between problems (6.129) and (6.132), and $\bar{x}$ and $\bar{\zeta}$ are optimal solutions of (6.129) and (6.132), respectively.

Proposition 6.32. *Suppose that mapping $F : \mathbb{R}^n \to \mathcal{Z}$ is convex and risk measure $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is proper and lower semicontinuous and satisfies conditions* (R1)–(R2). *Then $(\bar{x}, \bar{\zeta}) \in X \times \mathfrak{A}$ is a saddle point of $\Phi(x, \zeta)$ iff $\bar{\zeta} \in \partial\rho(\bar{Z})$ and*

$$0 \in \mathcal{N}_X(\bar{x}) + \mathbb{E}_{\bar{\zeta}}[\partial f_\omega(\bar{x})], \tag{6.133}$$

where $\bar{Z} := F(\bar{x})$.

Proof. By the definition, $(\bar{x}, \bar{\zeta})$ is a saddle point of $\Phi(x, \zeta)$ iff

$$\bar{x} \in \arg\min_{x \in X} \Phi(x, \bar{\zeta}) \quad \text{and} \quad \bar{\zeta} \in \arg\max_{\zeta \in \mathfrak{A}} \Phi(\bar{x}, \zeta). \tag{6.134}$$

The first of the above conditions means that $\bar{x} \in \arg\min_{x \in X} \psi(x)$, where

$$\psi(x) := \int_\Omega f(x, \omega)\bar{\zeta}(\omega) dP(\omega).$$

Since X is convex and $\psi(\cdot)$ is convex real valued, by the standard optimality conditions this holds iff $0 \in \mathcal{N}_X(\bar{x}) + \partial\psi(\bar{x})$. Moreover, by Theorem 7.47 we have $\partial\psi(\bar{x}) = \mathbb{E}_{\bar{\zeta}}[\partial f_\omega(\bar{x})]$. Therefore, condition (6.133) and the first condition in (6.134) are equivalent. The second condition (6.134) and the condition $\bar{\zeta} \in \partial\rho(\bar{Z})$ are equivalent by (6.42). $\square$

Under the assumptions of Proposition 6.32, existence of $\bar{\zeta} \in \partial\rho(\bar{Z})$ in (6.133) can be viewed as an optimality condition for problem (6.128). Sufficiency of that condition follows directly from the fact that it implies that $(\bar{x}, \bar{\zeta})$ is a saddle point of the min-max problem (6.129). In order for that condition to be necessary we need to verify existence of a saddle point for problem (6.129).

Proposition 6.33. *Let $\bar{x}$ be an optimal solution of the problem* (6.128). *Suppose that the mapping $F : \mathbb{R}^n \to \mathcal{Z}$ is convex and risk measure $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is proper and lower semicontinuous and satisfies conditions* (R1)–(R2) *and is continuous at $\bar{Z} := F(\bar{x})$. Then there exists $\bar{\zeta} \in \partial\rho(\bar{Z})$ such that $(\bar{x}, \bar{\zeta})$ is a saddle point of $\Phi(x, \zeta)$.*

Proof. By monotonicity of ρ (condition (R2)) it follows from the optimality of $\bar{x}$ that $(\bar{x}, \bar{Z})$ is an optimal solution of the problem

$$\operatorname*{Min}_{(x,Z)\in S} \rho(Z), \tag{6.135}$$

where $S := \big\{(x, Z) \in X \times \mathcal{Z} : F(x) \succeq Z\big\}$. Since F is convex, the set S is convex, and since F is continuous (see Lemma 6.9), the set S is closed. Also because ρ is convex and continuous at $\bar{Z}$, the following (first order) optimality condition holds at $(\bar{x}, \bar{Z})$ (see Remark 34, page 403):

$$0 \in \partial\rho(\bar{Z}) \times \{0\} + \mathcal{N}_S(\bar{x}, \bar{Z}). \tag{6.136}$$

This means that there exists $\bar{\zeta} \in \partial\rho(\bar{Z})$ such that $(-\bar{\zeta}, 0) \in \mathcal{N}_S(\bar{x}, \bar{Z})$. This in turn implies that

$$\langle \bar{\zeta}, Z - \bar{Z}\rangle \geq 0, \quad \forall (x, Z) \in S. \tag{6.137}$$

Setting $Z := F(x)$ we obtain that

$$\langle \bar{\zeta}, F(x) - F(\bar{x})\rangle \geq 0, \quad \forall x \in X. \tag{6.138}$$

It follows that $\bar{x}$ is a minimizer of $\langle \bar{\zeta}, F(x)\rangle$ over $x \in X$, and hence $\bar{x}$ is a minimizer of $\Phi(x, \bar{\zeta})$ over $x \in X$. That is, $\bar{x}$ satisfies first of the two conditions in (6.134). Moreover, as it was shown in the proof of Proposition 6.32, this implies condition (6.133), and hence $(\bar{x}, \bar{\zeta})$ is a saddle point by Proposition 6.32. □

Corollary 6.34. *Suppose that problem* (6.128) *has optimal solution $\bar{x}$, the mapping $F : \mathbb{R}^n \to \mathcal{Z}$ is convex and risk measure $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is proper and lower semicontinuous and satisfies conditions* (R1)–(R2)*, and is continuous at $\bar{Z} := F(\bar{x})$. Then there is no duality gap between problems* (6.129) *and* (6.132)*, and problem* (6.132) *has an optimal solution.*

Propositions 6.32 and 6.33 imply the following optimality conditions.

Theorem 6.35. *Suppose that mapping $F : \mathbb{R}^n \to \mathcal{Z}$ is convex and risk measure $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is proper and lower semicontinuous and satisfies conditions* (R1)–(R2)*. Consider a point $\bar{x} \in X$ and let $\bar{Z} := F(\bar{x})$. Then a sufficient condition for $\bar{x}$ to be an optimal solution of the problem* (6.128) *is existence of $\bar{\zeta} \in \partial\rho(\bar{Z})$ such that* (6.133) *holds. This condition is also necessary if ρ is continuous at $\bar{Z}$.*

It could be noted that if $\rho(\cdot) := \mathbb{E}[\cdot]$, then its subdifferential consists of unique subgradient $\bar{\zeta}(\cdot) \equiv 1$. In that case condition (6.133) takes the form

$$0 \in \mathcal{N}_X(\bar{x}) + \mathbb{E}[\partial f_\omega(\bar{x})]. \tag{6.139}$$

Note that since it is assumed that $F(x) \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$, the expectation $\mathbb{E}[f_\omega(x)]$ is well defined and finite valued for all x, and hence $\partial\mathbb{E}[f_\omega(x)] = \mathbb{E}[\partial f_\omega(x)]$ (see Theorem 7.47).

6.4.1 Dualization of Nonanticipativity Constraints

We assume again that $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and $\mathcal{Z}^* = \mathcal{L}_q(\Omega, \mathcal{F}, P)$, that $F : \mathbb{R}^n \to \mathcal{Z}$ is convex and $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is proper lower semicontinuous and satisfies conditions (R1) and (R2). A way to represent problem (6.128) is to consider the decision vector x as a function of the elementary event $\omega \in \Omega$ and then to impose an appropriate nonaniticipativity constraint. That is, let $\mathfrak{M}$ be a linear space of $\mathcal{F}$-measurable mappings $\chi : \Omega \to \mathbb{R}^n$. Define $F_\chi(\omega) := f(\chi(\omega), \omega)$ and

$$\mathfrak{M}_X := \{\chi \in \mathfrak{M} : \chi(\omega) \in X, \text{ a.e. } \omega \in \Omega\}. \tag{6.140}$$

We assume that the space $\mathfrak{M}$ is chosen in such a way that $F_\chi \in \mathcal{Z}$ for every $\chi \in \mathfrak{M}$ and for every $x \in \mathbb{R}^n$ the constant mapping $\chi(\omega) \equiv x$ belongs to $\mathfrak{M}$. Then we can write problem (6.128) in the following equivalent form:

$$\operatorname*{Min}_{(\chi,x)\in\mathfrak{M}_X\times\mathbb{R}^n} \rho(F_\chi) \text{ s.t. } \chi(\omega) = x, \text{ a.e. } \omega \in \Omega. \tag{6.141}$$

Formulation (6.141) allows developing a duality framework associated with the *nonanticipativity* constraint $\chi(\cdot) = x$. In order to formulate such duality, we need to specify the space $\mathfrak{M}$ and its dual. It looks natural to use $\mathfrak{M} := \mathcal{L}_{p'}(\Omega, \mathcal{F}, P; \mathbb{R}^n)$, for some $p' \in [1, +\infty)$, and its dual $\mathfrak{M}^* := \mathcal{L}_{q'}(\Omega, \mathcal{F}, P; \mathbb{R}^n)$, $q' \in (1, +\infty]$. It is also possible to employ $\mathfrak{M} := \mathcal{L}_\infty(\Omega, \mathcal{F}, P; \mathbb{R}^n)$. Unfortunately, this Banach space is not reflexive. Nevertheless, it can be paired with the space $\mathcal{L}_1(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ by defining the corresponding scalar product in the usual way. As long as the risk measure is lower semicontinuous and subdifferentiable in the corresponding weak topology, we can use this setting as well.

The (Lagrangian) dual of problem (6.141) can be written in the form

$$\operatorname*{Max}_{\lambda\in\mathfrak{M}^*} \left\{ \inf_{(\chi,x)\in\mathfrak{M}_X\times\mathbb{R}^n} L(\chi, x, \lambda) \right\}, \tag{6.142}$$

where

$$L(\chi, x, \lambda) := \rho(F_\chi) + \mathbb{E}\big[\lambda^{\mathsf{T}}(\chi - x)\big], \ (\chi, x, \lambda) \in \mathfrak{M} \times \mathbb{R}^n \times \mathfrak{M}^*. \tag{6.143}$$

Note that

$$\inf_{x\in\mathbb{R}^n} L(\chi, x, \lambda) = \begin{cases} L(\chi, 0, \lambda) & \text{if } \ \mathbb{E}[\lambda] = 0, \\ -\infty & \text{if } \ \mathbb{E}[\lambda] \neq 0. \end{cases}$$

Therefore the dual problem (6.143) can be rewritten in the form

$$\operatorname*{Max}_{\lambda\in\mathfrak{M}^*} \left\{ \inf_{\chi\in\mathfrak{M}_X} L_0(\chi, \lambda) \right\} \ \text{ s.t. } \mathbb{E}[\lambda] = 0, \tag{6.144}$$

where $L_0(\chi, \lambda) := L(\chi, 0, \lambda) = \rho(F_\chi) + \mathbb{E}[\lambda^{\mathsf{T}}\chi]$.

We have that the optimal value of problem (6.141) (which is the same as the optimal value of problem (6.128)) is greater than or equal to the optimal value of its dual (6.144). Moreover, under some regularity conditions, their optimal values are equal to each other. In particular, if Lagrangian $L(\chi, x, \lambda)$ has a saddle point $((\bar{\chi}, \bar{x}), \bar{\lambda})$, then there is no duality gap between problems (6.141) and (6.144), and $(\bar{\chi}, \bar{x})$ and $\bar{\lambda}$ are optimal solutions of problems

(6.141) and (6.144), respectively. Noting that $L(\chi, 0, \lambda)$ is linear in x and in λ, we have that $((\bar{\chi}, \bar{x}), \bar{\lambda})$ is a saddle point of $L(\chi, x, \lambda)$ iff the following conditions hold:

$$\begin{aligned} &\bar{\chi}(\omega) = \bar{x}, \text{ a.e. } \omega \in \Omega, \quad \text{and } \mathbb{E}[\bar{\lambda}] = 0, \\ &\bar{\chi} \in \arg \min_{\chi \in \mathfrak{M}_X} L_0(\chi, \bar{\lambda}). \end{aligned} \tag{6.145}$$

Unfortunately, it may be not be easy to verify existence of such saddle point.

We can approach the duality analysis by conjugate duality techniques. For a perturbation vector $y \in \mathfrak{M}$ consider the problem

$$\operatorname*{Min}_{(\chi,x)\in\mathfrak{M}_X\times\mathbb{R}^n} \rho(F_\chi) \text{ s.t. } \chi(\omega) = x + y(\omega), \tag{6.146}$$

and let $\vartheta(y)$ be its optimal value. Note that a perturbation in the vector x, in the constraints of problem (6.141), can be absorbed into $y(\omega)$. Clearly for $y = 0$, problem (6.146) coincides with the unperturbed problem (6.141), and $\vartheta(0)$ is the optimal value of the unperturbed problem (6.141). Assume that $\vartheta(0)$ is finite. Then there is no duality gap between problem (6.141) and its dual (6.142) iff $\vartheta(y)$ is lower semicontinuous at $y = 0$. Again it may be not easy to verify lower semicontinuity of the optimal value function $\vartheta : \mathfrak{M} \to \mathbb{R}$. By the general theory of conjugate duality we have the following result.

Proposition 6.36. *Suppose that $F : \mathbb{R}^n \to \mathcal{Z}$ is convex, $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ satisfies conditions* (R1)–(R2) *and the function $\rho(F_\chi)$, from $\mathfrak{M}$ to $\overline{\mathbb{R}}$, is lower semicontinuous. Suppose, further, that $\vartheta(0)$ is finite and $\vartheta(y) < +\infty$ for all y in a neighborhood (in the norm topology) of $0 \in \mathfrak{M}$. Then there is no duality gap between problems* (6.141) *and* (6.142)*, and the dual problem* (6.142) *has an optimal solution.*

Proof. Since ρ satisfies conditions (R1) and (R2) and F is convex, we have that the function $\rho(F_\chi)$ is convex, and by the assumption it is lower semicontinuous. The assertion then follows by a general result of conjugate duality for Banach spaces (see Theorem 7.77). □

In order to apply the above result, we need to verify lower semicontinuity of the function $\rho(F_\chi)$. This function is lower semicontinuous if $\rho(\cdot)$ is lower semicontinuous and the mapping $\chi \mapsto F_\chi$, from $\mathfrak{M}$ to $\mathcal{Z}$, is continuous. If the set Ω is finite, and hence the spaces $\mathcal{Z}$ and $\mathfrak{M}$ are finite dimensional, then continuity of $\chi \mapsto F_\chi$ follows from the continuity of F. In the infinite dimensional setting this should be verified by specialized methods. The assumption that $\vartheta(0)$ is finite means that the optimal value of the problem (6.141) is finite, and the assumption that $\vartheta(y) < +\infty$ means that the corresponding problem (6.146) has a feasible solution.

Interchangeability Principle for Risk Measures

By removing the nonanticipativity constraint $\chi(\cdot) = x$, we obtain the following relaxation of the problem (6.141):

$$\operatorname*{Min}_{\chi\in\mathfrak{M}_X} \rho(F_\chi), \tag{6.147}$$

where $\mathfrak{M}_X$ is defined in (6.140). Similarly to the interchangeability principle for the expectation operator (Theorem 7.80), we have the following result for monotone risk measures. By $\inf_{x\in X} F(x)$ we denote the pointwise minimum, i.e.,

$$\left[\inf_{x\in X} F(x)\right](\omega) := \inf_{x\in X} f(x,\omega), \quad \omega\in\Omega. \tag{6.148}$$

Proposition 6.37. *Let $\mathcal{Z} := \mathcal{L}_p(\Omega,\mathcal{F},P)$ and $\mathfrak{M} := \mathcal{L}_{p'}(\Omega,\mathcal{F},P;\mathbb{R}^n)$, where $p,p'\in[1,+\infty]$, $\mathfrak{M}_X$ be defined in* (6.140)*, $\rho:\mathcal{Z}\to\overline{\mathbb{R}}$ be a proper risk measure satisfying monotonicity condition* (R2)*, and $F:\mathbb{R}^n\to\mathcal{Z}$ be such that $\inf_{x\in X} F(x)\in\mathcal{Z}$. Suppose that ρ is continuous at $\Psi := \inf_{x\in X} F(x)$. Then*

$$\inf_{\chi\in\mathfrak{M}_X} \rho(F_\chi) = \rho\left(\inf_{x\in X} F(x)\right). \tag{6.149}$$

Proof. For any $\chi\in\mathfrak{M}_X$ we have that $\chi(\cdot)\in X$, and hence the following inequality holds:

$$\left[\inf_{x\in X} F(x)\right](\omega) \le F_\chi(\omega) \ \text{ a.e. } \omega\in\Omega.$$

By monotonicity of ρ this implies that $\rho(\Psi)\le\rho(F_\chi)$, and hence

$$\rho(\Psi) \le \inf_{\chi\in\mathfrak{M}_X} \rho(F_\chi). \tag{6.150}$$

Since ρ is proper we have that $\rho(\Psi) > -\infty$. If $\rho(\Psi)=+\infty$, then by (6.150) the left-hand side of (6.149) is also $+\infty$ and hence (6.149) holds. Therefore we can assume that $\rho(\Psi)$ is finite.

Let us derive now the converse of (6.150) inequality. Since it is assumed that $\Psi\in\mathcal{Z}$, we have that $\Psi(\omega)$ is finite valued for a.e. $\omega\in\Omega$ and measurable. Therefore, for a sequence $\varepsilon_k\downarrow 0$ and a.e. $\omega\in\Omega$ and all $k\in\mathbb{N}$, we can choose $\chi_k(\omega)\in X$ such that $|f(\chi_k(\omega),\omega)-\Psi(\omega)|\le\varepsilon_k$ and $\chi_k(\cdot)$ are measurable. We also can truncate $\chi_k(\cdot)$, if necessary, in such a way that each χ_k belongs to $\mathfrak{M}_X$, and $f(\chi_k(\omega),\omega)$ monotonically converges to $\Psi(\omega)$ for a.e. $\omega\in\Omega$. We have then that $f(\chi_k(\cdot),\cdot)-\Psi(\cdot)$ is nonnegative valued and is dominated by a function from the space $\mathcal{Z}$. It follows by the Lebesgue dominated convergence theorem that F_{χ_k} converges to Ψ in the norm topology of $\mathcal{Z}$. Since ρ is continuous at Ψ, it follows that $\rho(F_{\chi_k})$ tends to $\rho(\Psi)$. Also $\inf_{\chi\in\mathfrak{M}_X}\rho(F_\chi)\le\rho(F_{\chi_k})$, and hence the required converse inequality

$$\inf_{\chi\in\mathfrak{M}_X} \rho(F_\chi) \le \rho(\Psi) \tag{6.151}$$

follows. ◻

Remark 22. It follows from (6.149) that if

$$\bar\chi \in \arg\min_{\chi\in\mathfrak{M}_X} \rho(F_\chi), \tag{6.152}$$

then

$$\bar\chi(\omega) \in \arg\min_{x\in X} f(x,\omega) \ \text{ a.e. } \omega\in\Omega. \tag{6.153}$$

Conversely, suppose that the function $f(x, \omega)$ is random lower semicontinuous. Then the multifunction $\omega \mapsto \arg\min_{x\in X} f(x, \omega)$ is measurable. Therefore, $\bar{\chi}(\omega)$ in the left-hand side of (6.153) can be chosen to be measurable. If, moreover, $\bar{\chi} \in \mathfrak{M}$ (this holds, in particular, if the set X is bounded and hence $\bar{\chi}(\cdot)$ is bounded), then the inclusion (6.152) follows.

Consider now a setting of two-stage programming. That is, suppose that the function $[F(x)](\omega) = f(x, \omega)$ of the first-stage problem

$$\operatorname*{Min}_{x\in X} \rho(F(x)) \tag{6.154}$$

is given by the optimal value of the second-stage problem

$$\operatorname*{Min}_{y\in \mathcal{G}(x,\omega)} g(x, y, \omega), \tag{6.155}$$

where $g : \mathbb{R}^n \times \mathbb{R}^m \times \Omega \to \mathbb{R}$ and $\mathcal{G} : \mathbb{R}^n \times \Omega \rightrightarrows \mathbb{R}^m$. Under appropriate regularity conditions, from which the most important is the monotonicity condition (R2), we can apply the interchangeability principle to the optimization problem (6.155) to obtain

$$\rho(F(x)) = \inf_{y(\cdot)\in \mathcal{G}(x,\cdot)} \rho(g(x, y(\omega), \omega)), \tag{6.156}$$

where now $y(\cdot)$ is an element of an appropriate functional space and the notation $y(\cdot) \in \mathcal{G}(x, \cdot)$ means that $y(\omega) \in \mathcal{G}(x, \omega)$ w.p. 1. If the interchangeability principle (6.156) holds, then the two-stage problem (6.154)–(6.155) can be written as one large optimization problem:

$$\operatorname*{Min}_{x\in X,\; y(\cdot)\in \mathcal{G}(x,\cdot)} \rho(g(x, y(\omega), \omega)). \tag{6.157}$$

In particular, suppose that the set Ω is finite, say $\Omega = \{\omega_1, \dots, \omega_K\}$, i.e., there is a finite number K of scenarios. In that case we can view function $Z : \Omega \to \mathbb{R}$ as vector $(Z(\omega_1), \dots, Z(\omega_K)) \in \mathbb{R}^K$ and hence identify the space $\mathcal{Z}$ with $\mathbb{R}^K$. Then problem (6.157) takes the form

$$\operatorname*{Min}_{x\in X,\; y_k\in \mathcal{G}(x,\omega_k),\, k=1,\dots,K} \rho\left[(g(x, y_1, \omega_1), \dots, g(x, y_K, \omega_K))\right]. \tag{6.158}$$

Moreover, consider the linear case where $X := \{x : Ax = b,\ x \ge 0\}$, $g(x, y, \omega) := c^{\mathsf T}x + q(\omega)^{\mathsf T}y$ and

$$\mathcal{G}(x, \omega) := \{y : T(\omega)x + W(\omega)y = h(\omega),\ y \ge 0\}.$$

Assume that ρ satisfies conditions (R1)–(R3) and the set $\Omega = \{\omega_1, \dots, \omega_K\}$ is finite. Then problem (6.158) takes the form

$$\begin{aligned} &\operatorname*{Min}_{x, y_1,\dots,y_K} c^{\mathsf T}x + \rho\left[\left(q_1^{\mathsf T}y_1, \dots, q_K^{\mathsf T}y_K\right)\right] \\ &\quad \text{s.t. } Ax = b,\ x \ge 0,\ T_k x + W_k y_k = h_k,\ y_k \ge 0,\ k = 1, \dots, K, \end{aligned} \tag{6.159}$$

where $(q_k, T_k, W_k, h_k) := (q(\omega_k), T(\omega_k), W(\omega_k), h(\omega_k))$, $k = 1, \dots, K$.

6.4.2 Examples

Let $\mathcal{Z} := \mathcal{L}_1(\Omega, \mathcal{F}, P)$ and consider

$$\rho(Z) := \mathbb{E}[Z] + \inf_{t \in \mathbb{R}} \mathbb{E}\big\{\beta_1[t - Z]_+ + \beta_2[Z - t]_+\big\}, \quad Z \in \mathcal{Z}, \tag{6.160}$$

where $\beta_1 \in [0, 1]$ and $\beta_2 \geq 0$ are some constants. Properties of this risk measure were studied in Example 6.16 (see (6.67) and (6.68) in particular). We can write the corresponding optimization problem (6.128) in the following equivalent form:

$$\operatorname*{Min}_{(x,t) \in X \times \mathbb{R}} \mathbb{E}\left\{f_\omega(x) + \beta_1[t - f_\omega(x)]_+ + \beta_2[f_\omega(x) - t]_+\right\}. \tag{6.161}$$

That is, by adding one extra variable we can formulate the corresponding optimization problem as an expectation minimization problem.

Risk Averse Optimization of an Inventory Model

Let us consider again the inventory model analyzed in section 1.2. Recall that the objective of that model is to minimize the total cost

$$F(x, d) = cx + b[d - x]_+ + h[x - d]_+,$$

where c, b, and h are nonnegative constants representing costs of ordering, backordering, and holding, respectively. Again we assume that $b > c > 0$, i.e., the backorder cost is *bigger* than the ordering cost. A risk averse extension of the corresponding (expected value) problem (1.4) can be formulated in the form

$$\operatorname*{Min}_{x \geq 0} \big\{f(x) := \rho[F(x, D)]\big\}, \tag{6.162}$$

where ρ is a specified risk measure.

Assume that the risk measure ρ is coherent, i.e., satisfies conditions (R1)–(R4), and that demand $D = D(\omega)$ belongs to an appropriate space $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$. Assume, further, that $\rho : \mathcal{Z} \to \mathbb{R}$ is *real valued*. It follows that there exists a convex set $\mathfrak{A} \subset \mathfrak{P}$, where $\mathfrak{P} \subset \mathcal{Z}^*$ is the set of probability density functions, such that

$$\rho(Z) = \sup_{\zeta \in \mathfrak{A}} \int_\Omega Z(\omega)\zeta(\omega)dP(\omega), \quad Z \in \mathcal{Z}.$$

Consequently we have that

$$\rho[F(x, D)] = \sup_{\zeta \in \mathfrak{A}} \int_\Omega F(x, D(\omega))\zeta(\omega)dP(\omega). \tag{6.163}$$

To each $\zeta \in \mathfrak{P}$ corresponds the cumulative distribution function H of D with respect to the measure $Q := \zeta dP$, that is,

$$H(z) = Q(D \leq z) = \mathbb{E}_\zeta[\mathbf{1}_{D \leq z}] = \int_{\{\omega : D(\omega) \leq z\}} \zeta(\omega)dP(\omega). \tag{6.164}$$

We have then that

$$\int_{\Omega} F(x, D(\omega))\zeta(\omega)dP(\omega) = \int F(x, z)dH(z).$$

Denote by $\mathfrak{M}$ the set of cumulative distribution functions H associated with densities $\zeta \in \mathfrak{A}$. The correspondence between $\zeta \in \mathfrak{A}$ and $H \in \mathfrak{M}$ is given by formula (6.164) and depends on $D(\cdot)$ and the reference probability measure P. Then we can rewrite (6.163) in the form

$$\rho[F(x, D)] = \sup_{H \in \mathfrak{M}} \int F(x, z)dH(z) = \sup_{H \in \mathfrak{M}} \mathbb{E}_H[F(x, D)]. \tag{6.165}$$

This leads to the following minimax formulation of the risk averse optimization problem (6.162):

$$\operatorname*{Min}_{x \geq 0} \sup_{H \in \mathfrak{M}} \mathbb{E}_H[F(x, D)]. \tag{6.166}$$

Note that we also have that $\rho(D) = \sup_{H \in \mathfrak{M}} \mathbb{E}_H[D]$.

In the subsequent analysis we deal with the minimax formulation (6.166), rather than the risk averse formulation (6.162), viewing $\mathfrak{M}$ as a given set of cumulative distribution functions. We show next that the minimax problem (6.166), and hence the risk averse problem (6.162), structurally is similar to the corresponding (expected value) problem (1.4). We assume that every $H \in \mathfrak{M}$ is such that $H(z) = 0$ for any $z < 0$. (Recall that the demand cannot be negative.) We also assume that $\sup_{H \in \mathfrak{M}} \mathbb{E}_H[D] < +\infty$, which follows from the assumption that $\rho(\cdot)$ is real valued.

Proposition 6.38. *Let $\mathfrak{M}$ be a set of cumulative distribution functions such that $H(z) = 0$ for any $H \in \mathfrak{M}$ and $z < 0$, and $\sup_{H \in \mathfrak{M}} \mathbb{E}_H[D] < +\infty$. Consider function $f(x) := \sup_{H \in \mathfrak{M}} \mathbb{E}_H[F(x, D)]$. Then there exists a cdf $\bar{H}$, depending on the set $\mathfrak{M}$ and $\eta := b/(b + h)$, such that $\bar{H}(z) = 0$ for any $z < 0$, and the function $f(x)$ can be written in the form*

$$f(x) = b \sup_{H \in \mathfrak{M}} \mathbb{E}_H[D] + (c - b)x + (b + h) \int_{-\infty}^{x} \bar{H}(z)dz. \tag{6.167}$$

Proof. We have (see formula (1.5)) that for $H \in \mathfrak{M}$,

$$\mathbb{E}_H[F(x, D)] = b\, \mathbb{E}_H[D] + (c - b)x + (b + h) \int_0^x H(z)dz.$$

Therefore we can write $f(x) = (c - b)x + (b + h)\mathfrak{g}(x)$, where

$$\mathfrak{g}(x) := \sup_{H \in \mathfrak{M}} \left\{ \eta\, \mathbb{E}_H[D] + \int_{-\infty}^{x} H(z)dz \right\}. \tag{6.168}$$

Since every $H \in \mathfrak{M}$ is a monotonically nondecreasing function, we have that $x \mapsto \int_{-\infty}^{x} H(z)dz$ is a convex function. It follows that the function $\mathfrak{g}(x)$ is given by the maximum of convex functions and hence is convex. Moreover, $\mathfrak{g}(x) \geq 0$ and

$$\mathfrak{g}(x) \leq \eta \sup_{H \in \mathfrak{M}} \mathbb{E}_H[D] + [x]_+, \tag{6.169}$$

and hence $\mathfrak{g}(x)$ is finite valued for any $x \in \mathbb{R}$. Also, for any $H \in \mathfrak{M}$ and $z < 0$ we have that $H(z) = 0$, and hence $\mathfrak{g}(x) = \eta \sup_{H\in\mathfrak{M}} \mathbb{E}_H[D]$ for any $x < 0$.

Consider the right-hand-side derivative of $\mathfrak{g}(x)$:

$$\mathfrak{g}^+(x) := \lim_{t\downarrow 0} \frac{\mathfrak{g}(x+t) - \mathfrak{g}(x)}{t},$$

and define $\bar{H}(\cdot) := \mathfrak{g}^+(\cdot)$. Since $\mathfrak{g}(x)$ is real valued convex, its right-hand-side derivative $\mathfrak{g}^+(x)$ exists and is finite, and for any $x \geq 0$ and $a < 0$,

$$\mathfrak{g}(x) = \mathfrak{g}(a) + \int_a^x \mathfrak{g}^+(z)dz = \eta \sup_{H\in\mathfrak{M}} \mathbb{E}_H[D] + \int_{-\infty}^x \bar{H}(z)dz. \tag{6.170}$$

Note that definition of the function $\mathfrak{g}(\cdot)$, and hence $\bar{H}(\cdot)$, involves the constant η and set $\mathfrak{M}$ only. Let us also observe that the right-hand-side derivative $\mathfrak{g}^+(x)$, of a real valued convex function, is monotonically nondecreasing and right-side continuous. Moreover, $\mathfrak{g}^+(x) = 0$ for $x < 0$ since $\mathfrak{g}(x)$ is constant for $x < 0$. We also have that $\mathfrak{g}^+(x)$ tends to one as $x \to +\infty$. Indeed, since $\mathfrak{g}^+(x)$ is monotonically nondecreasing it tends to a limit, denoted r, as $x \to +\infty$. We have then that $\mathfrak{g}(x)/x \to r$ as $x \to +\infty$. It follows from (6.169) that $r \leq 1$, and by (6.168) that for any $H \in \mathfrak{M}$,

$$\liminf_{x\to+\infty} \frac{\mathfrak{g}(x)}{x} \geq \liminf_{x\to+\infty} \frac{1}{x}\int_{-\infty}^x H(z)dz \geq 1,$$

and hence $r \geq 1$. It follows that $r = 1$.

We obtain that $\bar{H}(\cdot) = \mathfrak{g}^+(\cdot)$ is a cumulative distribution function of some probability distribution and the representation (6.167) holds. $\square$

It follows from the representation (6.167) that the set of optimal solutions of the risk averse problem (6.162) is an interval given by the set of κ-quantiles of the cdf $\bar{H}(\cdot)$, where $\kappa := \frac{b-c}{b+h}$. (Compare with Remark 1, page 3.)

In some specific cases it is possible to calculate the corresponding cdf $\bar{H}$ in a closed form. Consider the risk measure ρ defined in (6.160),

$$\rho(Z) := \mathbb{E}[Z] + \inf_{t\in\mathbb{R}} \mathbb{E}\big\{\beta_1[t-Z]_+ + \beta_2[Z-t]_+\big\},$$

where the expectations are taken with respect to some reference cdf $H^*(\cdot)$. The corresponding set $\mathfrak{M}$ is formed by cumulative distribution functions $H(\cdot)$ such that

$$(1-\beta_1)\int_S dH^* \leq \int_S dH \leq (1+\beta_2)\int_S dH^* \tag{6.171}$$

for any Borel set $S \subset \mathbb{R}$. (Compare with formula (6.69).) Recall that for $\beta_1 = 1$ this risk measure is $\rho(Z) = \mathsf{AV@R}_\alpha(Z)$ with $\alpha = 1/(1+\beta_2)$. Suppose that the reference distribution of the demand is uniform on the interval $[0, 1]$, i.e., $H^*(z) = z$ for $z \in [0, 1]$. It follows that any $H \in \mathfrak{M}$ is continuous, $H(0) = 0$ and $H(1) = 1$, and

$$\mathbb{E}_H[D] = \int_0^1 z dH(z) = zH(z)\big|_0^1 - \int_0^1 H(z)dz = 1 - \int_0^1 H(z)dz.$$

Consequently we can write function $\mathfrak{g}(x)$, defined in (6.168), for $x \in [0, 1]$ in the form

$$\mathfrak{g}(x) = \eta + \sup_{H \in \mathfrak{M}} \left\{ (1-\eta) \int_0^x H(z)dz - \eta \int_x^1 H(z)dz \right\}. \tag{6.172}$$

Suppose, further, that $h = 0$ (i.e., there are no holding costs) and hence $\eta = 1$. In that case

$$\mathfrak{g}(x) = 1 - \inf_{H \in \mathfrak{M}} \int_x^1 H(z)dz \quad \text{for } x \in [0, 1]. \tag{6.173}$$

By using the first inequality of (6.171) with $S := [0, z]$ we obtain that $H(z) \geq (1-\beta_1)z$ for any $H \in \mathfrak{M}$ and $z \in [0, 1]$. Similarly, by the second inequality of (6.171) with $S := [z, 1]$ we have that $H(z) \geq 1 + (1 + \beta_2)(z - 1)$ for any $H \in \mathfrak{M}$ and $z \in [0, 1]$. Consequently, the cdf

$$\bar{H}(z) := \max\{(1 - \beta_1)z, (1 + \beta_2)z - \beta_2\}, \quad z \in [0, 1], \tag{6.174}$$

is dominated by any other cdf $H \in \mathfrak{M}$, and it can be verified that $\bar{H} \in \mathfrak{M}$. Therefore, the minimum on the right-hand side of (6.173) is attained at $\bar{H}$ for any $x \in [0, 1]$, and hence this cdf $\bar{H}$ fulfills (6.167).

Note that for any $\beta_1 \in (0, 1)$ and $\beta_2 > 0$, the cdf $\bar{H}(\cdot)$ defined in (6.174) is strictly less than the reference cdf $H^*(\cdot)$ on the interval $(0, 1)$. Consequently, the corresponding risk averse optimal solution $\bar{H}^{-1}(\kappa)$ is bigger than the risk neutral optimal solution $H^{*-1}(\kappa)$. It should be not surprising that in the absence of holding costs it will be safer to order a larger quantity of the product.

Risk Averse Portfolio Selection

Consider the portfolio selection problem introduced in section 1.4. A risk averse formulation of the corresponding optimization problem can be written in the form

$$\operatorname*{Min}_{x \in X} \rho\big(-\textstyle\sum_{i=1}^n \xi_i x_i\big), \tag{6.175}$$

where ρ is a chosen risk measure and $X := \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = W_0, \; x \geq 0\}$. We use the negative of the return as an argument of the risk measure, because we developed our theory for the minimization, rather than maximization framework. An example below shows a possible problem with using risk measures with dispersions measured by variance or standard deviation.

Example 6.39. Let $n = 2$, $W_0 = 1$ and the risk measure ρ be of the form

$$\rho(Z) := \mathbb{E}[Z] + c\,\mathbb{D}[Z], \tag{6.176}$$

where $c > 0$ and $\mathbb{D}[\cdot]$ is a dispersion measure. Let the dispersion measure be either $\mathbb{D}[Z] := \sqrt{\mathbb{V}\mathrm{ar}[Z]}$ or $\mathbb{D}[Z] := \mathbb{V}\mathrm{ar}[Z]$. Suppose, further, that the space $\Omega := \{\omega_1, \omega_2\}$ consists of two points with associated probabilities p and $1 - p$ for some $p \in (0, 1)$. Define (random) return rates $\xi_1, \xi_2 : \Omega \to \mathbb{R}$ as follows: $\xi_1(\omega_1) = a$ and $\xi_1(\omega_2) = 0$, where a is some positive number, and $\xi_2(\omega_1) = \xi_2(\omega_2) = 0$. Obviously, it is better to invest in asset 1 than asset 2. Now, for $\mathbb{D}[Z] := \sqrt{\mathbb{V}\mathrm{ar}[Z]}$, we have that $\rho(-\xi_2) = 0$ and $\rho(-\xi_1) = -pa + ca\sqrt{p(1-p)}$. It follows that $\rho(-\xi_1) > \rho(-\xi_2)$ for any $c > 0$ and $p < (1 + c^{-2})^{-1}$. Similarly, for $\mathbb{D}[Z] := \mathbb{V}\mathrm{ar}[Z]$ we have that $\rho(-\xi_1) = -pa + ca^2 p(1-p)$, $\rho(-\xi_2) = 0$, and hence $\rho(-\xi_1) > \rho(-\xi_2)$ again, provided $p < 1 - (ca)^{-1}$. That is, although ξ_2 dominates ξ_1 in the sense that $\xi_1(\omega) \geq \xi_2(\omega)$ for *every* possible realization of $(\xi_1(\omega), \xi_2(\omega))$, we have that $\rho(\xi_1) > \rho(\xi_2)$.

Here $[F(x)](\omega) := -\xi_1(\omega)x_1 - \xi_2(\omega)x_2$. Let $\bar{x} := (1, 0)$ and $x^* := (0, 1)$. Note that the feasible set X is formed by vectors $t\bar{x} + (1-t)x^*$, $t \in [0, 1]$. We have that $[F(x)](\omega) = -\xi_1(\omega)x_1$, and hence $[F(\bar{x})](\omega)$ is dominated by $[F(x)](\omega)$ for any $x \in X$ and $\omega \in \Omega$. And yet, under the specified conditions, we have that $\rho[F(\bar{x})] = \rho(-\xi_1)$ is greater than $\rho[F(x^*)] = \rho(-\xi_2)$, and hence $\bar{x}$ is *not* an optimal solution of the corresponding optimization (minimization) problem. This should be not surprising, because the chosen risk measure is not monotone, i.e., it does not satisfy the condition (R2), for $c > 0$. (See Examples 6.18 and 6.19.) ■

Suppose now that ρ is a real valued coherent risk measure. We can then write problem (6.175) in the corresponding min-max form (6.131), that is,

$$\underset{x \in X}{\mathrm{Min}} \sup_{\zeta \in \mathfrak{A}} \sum_{i=1}^{n} \left(-\mathbb{E}_\zeta[\xi_i]\right) x_i .$$

Equivalently,

$$\underset{x \in X}{\mathrm{Max}} \inf_{\zeta \in \mathfrak{A}} \sum_{i=1}^{n} \left(\mathbb{E}_\zeta[\xi_i]\right) x_i . \tag{6.177}$$

Since the feasible set X is compact, problem (6.175) always has an optimal solution $\bar{x}$. Also (see Proposition 6.33), the min-max problem (6.177) has a saddle point, and $(\bar{x}, \bar{\zeta})$ is a saddle point iff

$$\bar{\zeta} \in \partial\rho(\bar{Z}) \text{ and } \bar{x} \in \arg\max_{x \in X} \sum_{i=1}^{n} \bar{\mu}_i x_i , \tag{6.178}$$

where $\bar{Z}(\omega) := -\sum_{i=1}^{n} \xi_i(\omega)\bar{x}_i$ and $\bar{\mu}_i := \mathbb{E}_{\bar{\zeta}}[\xi_i]$.

An interesting insight into the risk averse solution is provided by its game-theoretical interpretation. For $W_0 = 1$ the portfolio allocations x can be interpreted as a *mixed strategy* of the investor. (For another W_0, the fractions x_i / W_0 are the mixed strategy.) The measure ζ represents the mixed strategy of the opponent (the market). It is chosen not from the set of all possible mixed strategies but rather from the set $\mathfrak{A}$. The risk averse solution (6.178) corresponds to the equilibrium of the game.

It is not difficult to see that the set $\arg\max_{x \in X} \sum_{i=1}^{n} \bar{\mu}_i x_i$ is formed by all convex combinations of vectors $W_0 e_i$, $i \in \mathcal{I}$, where $e_i \in \mathbb{R}^n$ denotes the ith coordinate vector (with zero entries except the ith entry equal to 1), and

$$\mathcal{I} := \left\{ i' : \bar{\mu}_{i'} = \max\nolimits_{1 \leq i \leq n} \bar{\mu}_i ,\ i' = 1, \ldots, n \right\} .$$

Also $\partial\rho(Z) \subset \mathfrak{A}$; see formula (6.43) for the subdifferential $\partial\rho(Z)$.

6.5 Statistical Properties of Risk Measures

All examples of risk measures discussed in section 6.3.2 were constructed with respect to a reference probability measure (distribution) P. Suppose now that the "true" probability distribution P is estimated by an empirical measure (distribution) P_N based on a sample of size N. In this section we discuss statistical properties of the respective estimates of the "true values" of the corresponding risk measures.

6.5.1 Average Value-at-Risk

Recall that the Average Value-at-Risk, $\mathsf{AV@R}_\alpha(Z)$, at a level $\alpha \in (0,1)$ of a random variable Z, is given by the optimal value of the minimization problem

$$\operatorname*{Min}_{t\in\mathbb{R}} \mathbb{E}\left\{t + \alpha^{-1}[Z-t]_+\right\}, \tag{6.179}$$

where the expectation is taken with respect to the probability distribution P of Z. We assume that $\mathbb{E}|Z| < +\infty$, which implies that $\mathsf{AV@R}_\alpha(Z)$ is finite. Suppose now that we have an iid random sample $Z^1, \dots, Z^N$ of N realizations of Z. Then we can estimate $\theta^* := \mathsf{AV@R}_\alpha(Z)$ by replacing distribution P with its empirical estimate[48] $P_N := \frac{1}{N}\sum_{j=1}^N \Delta(Z^j)$. This leads to the sample estimate $\hat\theta_N$, of $\theta^* = \mathsf{AV@R}_\alpha(Z)$, given by the optimal value of the following problem:

$$\operatorname*{Min}_{t\in\mathbb{R}} \left\{ t + \frac{1}{\alpha N}\sum_{j=1}^N \left[Z^j - t\right]_+ \right\}. \tag{6.180}$$

Let us observe that problem (6.179) can be viewed as a stochastic programming problem and problem (6.180) as its sample average approximation. That is,

$$\theta^* = \inf_{t\in\mathbb{R}} f(t) \quad \text{and} \quad \hat\theta_N = \inf_{t\in\mathbb{R}} \hat f_N(t),$$

where

$$f(t) = t + \alpha^{-1}\mathbb{E}[Z-t]_+ \quad \text{and} \quad \hat f_N(t) = t + \frac{1}{\alpha N}\sum_{j=1}^N \left[Z^j - t\right]_+.$$

Therefore, results of section 5.1 can be applied here in a straightforward way. Recall that the set of optimal solutions of problem (6.179) is the interval $[t^*, t^{**}]$, where

$$t^* = \inf\{z : H_Z(z) \ge 1-\alpha\} = \mathsf{V@R}_\alpha(Z) \quad \text{and} \quad t^{**} = \sup\{z : H_Z(z) \le 1-\alpha\}$$

are the respective left- and right-side $(1-\alpha)$-quantiles of the distribution of Z (see page 258). Since for any $\alpha \in (0,1)$ the interval $[t^*, t^{**}]$ is finite and problem (6.179) is convex, we have by Theorem 5.4 that

$$\hat\theta_N \to \theta^* \ \text{ w.p. 1 as } N \to \infty. \tag{6.181}$$

That is, $\hat\theta_N$ is a consistent estimator of $\theta^* = \mathsf{AV@R}_\alpha(Z)$.

[48]Recall that $\Delta(z)$ denotes measure of mass one at point z.

Assume now that $\mathbb{E}[Z^2] < +\infty$. Then the assumptions (A1) and (A2) of Theorem 5.7 hold, and hence

$$\hat{\theta}_N = \inf_{t \in [t^*, t^{**}]} \hat{f}_N(t) + o_p(N^{-1/2}). \tag{6.182}$$

Moreover, if $t^* = t^{**}$, i.e., the left- and right-side $(1-\alpha)$-quantiles of the distribution of Z are the same, then

$$N^{1/2}\left(\hat{\theta}_N - \theta^*\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2), \tag{6.183}$$

where $\sigma^2 = \alpha^{-2}\mathbb{V}\mathrm{ar}\,([Z - t^*]_+)$.

The estimator $\hat{\theta}_N$ has a negative bias, i.e., $\mathbb{E}[\hat{\theta}_N] - \theta^* \le 0$, and (see Proposition 5.6)

$$\mathbb{E}[\hat{\theta}_N] \le \mathbb{E}[\hat{\theta}_{N+1}], \quad N = 1, \dots, \tag{6.184}$$

i.e., the bias is monotonically decreasing with increase of the sample size N. If $t^* = t^{**}$, then this bias is of order $O(N^{-1})$ and can be estimated using results of section 5.1.3. The first and second order derivatives of the expectation function $f(t)$ here are $f'(t) = 1 + \alpha^{-1}(H_Z(t) - 1)$, provided that the cumulative distribution function $H_Z(\cdot)$ is continuous at t, and $f''(t) = \alpha^{-1}h_Z(t)$, provided that the density $h_Z(t) = \partial H_Z(t)/\partial t$ exists. We obtain (see Theorem 5.8 and the discussion on page 168), under appropriate regularity conditions, in particular if $t^* = t^{**} = \mathsf{V@R}_\alpha(Z)$ and the density $h_Z(t^*) = \partial H_Z(t^*)/\partial t$ exists and $h_Z(t^*) \ne 0$, that

$$\begin{aligned} \hat{\theta}_N - \hat{f}_N(t^*) &= N^{-1} \inf\nolimits_{\tau \in \mathbb{R}} \left\{ \tau Z + \tfrac{1}{2}\tau^2 f''(t^*) \right\} + o_p(N^{-1}) \\ &= -\tfrac{\alpha Z^2}{2N h_Z(t^*)} + o_p(N^{-1}), \end{aligned} \tag{6.185}$$

where $Z \sim \mathcal{N}(0, \gamma^2)$ with

$$\gamma^2 = \mathbb{V}\mathrm{ar}\left(\alpha^{-1}\frac{\partial [Z - t^*]_+}{\partial t}\right) = \frac{H_Z(t^*)(1 - H_Z(t^*))}{\alpha^2} = \frac{1-\alpha}{\alpha}.$$

Consequently, under appropriate regularity conditions,

$$N\left[\hat{\theta}_N - \hat{f}_N(t^*)\right] \xrightarrow{\mathcal{D}} -\left[\frac{1-\alpha}{2h_Z(t^*)}\right]\chi_1^2 \tag{6.186}$$

and (see Remark 32 on page 382)

$$\mathbb{E}[\hat{\theta}_N] - \theta^* = -\frac{1-\alpha}{2N h_Z(t^*)} + o(N^{-1}). \tag{6.187}$$

6.5.2 Absolute Semideviation Risk Measure

Consider the mean absolute semideviation risk measure

$$\rho_c(Z) := \mathbb{E}\left\{Z + c[Z - \mathbb{E}(Z)]_+\right\}, \tag{6.188}$$

where $c \in [0, 1]$ and the expectation is taken with respect to the probability distribution P of Z. We assume that $\mathbb{E}|Z| < +\infty$, and hence $\rho_c(Z)$ is finite. For a random sample

$Z^1, \ldots, Z^N$ of Z, the corresponding estimator of $\theta^* := \rho_c(Z)$ is

$$\hat{\theta}_N = N^{-1} \sum_{j=1}^{N} \left(Z^j + c[Z^j - \bar{Z}]_+ \right), \tag{6.189}$$

where $\bar{Z} = N^{-1} \sum_{j=1}^{N} Z^j$.

We have that $\rho_c(Z)$ is equal to the optimal value of the following convex–concave minimax problem

$$\operatorname*{Min}_{t \in \mathbb{R}} \max_{\gamma \in [0,1]} \mathbb{E}\left[F(t, \gamma, Z)\right], \tag{6.190}$$

where

$$\begin{aligned} F(t, \gamma, z) &:= z + c\gamma[z - t]_+ + c(1-\gamma)[t - z]_+ \\ &= z + c[z - t]_+ + c(1-\gamma)(z - t). \end{aligned} \tag{6.191}$$

This follows by virtue of Corollary 6.3. More directly we can argue as follows. Denote $\mu := \mathbb{E}[Z]$. We have that

$$\begin{aligned} \sup_{\gamma \in [0,1]} \quad & \mathbb{E}\left\{ Z + c\gamma[Z - t]_+ + c(1-\gamma)[t - Z]_+ \right\} \\ & = \mathbb{E}[Z] + c \max\left\{ \mathbb{E}([Z - t]_+), \mathbb{E}([t - Z]_+) \right\}. \end{aligned}$$

Moreover, $\mathbb{E}([Z - t]_+) = \mathbb{E}([t - Z]_+)$ if $t = \mu$, and either $\mathbb{E}([Z - t]_+)$ or $\mathbb{E}([t - Z]_+)$ is bigger than $\mathbb{E}([Z - \mu]_+)$ if $t \neq \mu$. This implies the assertion and also shows that the minimum in (6.190) is attained at unique point $t^* = \mu$. It also follows that the set of saddle points of the minimax problem (6.190) is given by $\{\mu\} \times [\gamma^*, \gamma^{**}]$, where

$$\gamma^* = \Pr(Z < \mu) \ \text{ and } \ \gamma^{**} = \Pr(Z \le \mu) = H_Z(\mu). \tag{6.192}$$

In particular, if the cdf $H_Z(\cdot)$ is continuous at $\mu = \mathbb{E}[Z]$, then there is unique saddle point $(\mu, H_Z(\mu))$.

Consequently, $\hat{\theta}_N$ is equal to the optimal value of the corresponding SAA problem

$$\operatorname*{Min}_{t \in \mathbb{R}} \max_{\gamma \in [0,1]} N^{-1} \sum_{j=1}^{N} F(t, \gamma, Z^j). \tag{6.193}$$

Therefore we can apply results of section 5.1.4 in a straightforward way. We obtain that $\hat{\theta}_N$ converges w.p. 1 to θ^* as $N \to \infty$. Moreover, assuming that $\mathbb{E}[Z^2] < +\infty$ we have by Theorem 5.10 that

$$\begin{aligned} \hat{\theta}_N &= \max_{\gamma \in [\gamma^*, \gamma^{**}]} N^{-1} \textstyle\sum_{j=1}^{N} F(\mu, \gamma, Z^j) + o_p(N^{-1/2}) \\ &= \bar{Z} + cN^{-1} \textstyle\sum_{j=1}^{N} [Z^j - \mu]_+ + c\Psi(\bar{Z} - \mu) + o_p(N^{-1/2}), \end{aligned} \tag{6.194}$$

where $\bar{Z} = N^{-1} \sum_{j=1}^{N} Z^j$ and function $\Psi(\cdot)$ is defined as

$$\Psi(z) := \begin{cases} (1-\gamma^*)z & \text{if} \quad z > 0, \\ (1-\gamma^{**})z & \text{if} \quad z \le 0. \end{cases}$$

If, moreover, the cdf $H_Z(\cdot)$ is continuous at μ, and hence $\gamma^* = \gamma^{**} = H_Z(\mu)$, then

$$N^{1/2}(\hat{\theta}_N - \theta^*) \xrightarrow{\mathcal{D}} N(0, \sigma^2), \tag{6.195}$$

where $\sigma^2 = \mathbb{V}\mathrm{ar}[F(\mu, H_Z(\mu), Z)]$.

This analysis can be extended to risk averse optimization problems of the form (6.128). That is, consider problem

$$\operatorname*{Min}_{x \in X} \Big\{ \rho_c[G(x,\xi)] = \mathbb{E}\big\{G(x,\xi) + c[G(x,\xi) - \mathbb{E}(G(x,\xi))]_+\big\}\Big\}, \tag{6.196}$$

where $X \subset \mathbb{R}^n$ and $G : X \times \Xi \to \mathbb{R}$. Its SAA is obtained by replacing the true distribution of the random vector ξ with the empirical distribution associated with a random sample $\xi^1, \dots, \xi^N$, that is,

$$\operatorname*{Min}_{x \in X} \frac{1}{N} \sum_{j=1}^{N} \left\{ G(x, \xi^j) + c\left[G(x,\xi^j) - \tfrac{1}{N}\textstyle\sum_{j=1}^{N} G(x,\xi^j)\right]_+ \right\}. \tag{6.197}$$

Assume that the set X is convex compact and function $G(\cdot, \xi)$ is convex for a.e. ξ. Then, for $c \in [0, 1]$, problems (6.196) and (6.197) are convex. By using the min-max representation (6.190), problem (6.196) can be written as the minimax problem

$$\operatorname*{Min}_{(x,t) \in X \times \mathbb{R}} \; \max_{\gamma \in [0,1]} \mathbb{E}\left[F(t, \gamma, G(x,\xi))\right], \tag{6.198}$$

where function $F(t, \gamma, z)$ is defined in (6.191). The function $F(t, \gamma, z)$ is convex and monotonically increasing in z. Therefore, by convexity of $G(\cdot, \xi)$, the function $F(t, \gamma, G(x, \xi))$ is convex in $x \in X$, and hence (6.198) is a convex–concave minimax problem. Consequently, results of section 5.1.4 can be applied.

Let ϑ^* and $\hat{\vartheta}_N$ be the optimal values of the true problem (6.196) and the SAA problem (6.197), respectively, and S be the set of optimal solutions of the true problem (6.196). By Theorem 5.10 and the above analysis we obtain, assuming that conditions specified in Theorem 5.10 are satisfied, that

$$\hat{\vartheta}_N = N^{-1} \inf_{\substack{x \in S \\ t = \mathbb{E}[G(x,\xi)]}} \; \max_{\gamma \in [\gamma^*, \gamma^{**}]} \left\{ \sum_{j=1}^{N} F\left(t, \gamma, G(x, \xi^j)\right) \right\} + o_p(N^{-1/2}), \tag{6.199}$$

where

$$\gamma^* := \Pr\big\{G(x,\xi) < \mathbb{E}[G(x,\xi)]\big\} \;\text{ and }\; \gamma^{**} := \Pr\big\{G(x,\xi) \le \mathbb{E}[G(x,\xi)]\big\}, \quad x \in S.$$

Note that the points $\big((x, \mathbb{E}[G(x,\xi)]), \gamma\big)$, where $x \in S$ and $\gamma \in [\gamma^*, \gamma^{**}]$, form the set of saddle points of the convex–concave minimax problem (6.198), and hence the interval $[\gamma^*, \gamma^{**}]$ is the same for any $x \in S$.

Moreover, assume that $S = \{\bar{x}\}$ is a singleton, i.e., problem (6.196) has unique optimal solution $\bar{x}$, and the cdf of the random variable $Z = G(\bar{x}, \xi)$ is continuous at $\mu := \mathbb{E}[G(\bar{x}, \xi)]$, and hence $\gamma^* = \gamma^{**}$. Then it follows that $N^{1/2}(\hat{\vartheta}_N - \vartheta^*)$ converges in distribution to normal with zero mean and variance

$$\mathbb{V}\mathrm{ar}\big\{G(\bar{x},\xi) + c[G(\bar{x},\xi) - \mu]_+ + c(1-\gamma^*)(G(\bar{x},\xi) - \mu)\big\}.$$

6.5.3 Von Mises Statistical Functionals

In the two examples, of $\mathsf{AV@R}_\alpha$ and absolute semideviation, of risk measures considered in the above sections it was possible to use their variational representations in order to apply results and methods developed in section 5.1. A possible approach to deriving large sample asymptotics of law invariant coherent risk measures is to use the Kusuoka representation described in Theorem 6.24 (such approach was developed in [147]). In this section we discuss an alternative approach of *Von Mises statistical functionals* borrowed from statistics. We view now a (law invariant) risk measure $\rho(Z)$ as a function $\mathfrak{F}(P)$ of the corresponding probability measure P. For example, with the (upper) semideviation risk measure $\sigma_p^+[Z]$, defined in (6.5), we associate the functional

$$\mathfrak{F}(P) := \left(\mathbb{E}_P\left[\left(Z - \mathbb{E}_P[Z]\right)_+^p\right]\right)^{1/p}. \tag{6.200}$$

The sample estimate of $\mathfrak{F}(P)$ is obtained by replacing probability measure P with the empirical measure P_N. That is, we estimate $\theta^* = \mathfrak{F}(P)$ by $\hat{\theta}_N = \mathfrak{F}(P_N)$.

Let Q be an arbitrary probability measure, defined on the same probability space as P, and consider the convex combination $(1-t)P + tQ = P + t(Q-P)$, with $t \in [0,1]$, of P and Q. Suppose that the following limit exists:

$$\mathfrak{F}'(P, Q-P) := \lim_{t \downarrow 0} \frac{\mathfrak{F}(P + t(Q-P)) - \mathfrak{F}(P)}{t}. \tag{6.201}$$

The above limit is just the directional derivative of $\mathfrak{F}(\cdot)$ at P in the direction $Q-P$. If, moreover, the directional derivative $\mathfrak{F}'(P,\cdot)$ is linear, then $\mathfrak{F}(\cdot)$ is Gâteaux differentiable at P. Consider now the approximation

$$\mathfrak{F}(P_N) - \mathfrak{F}(P) \approx \mathfrak{F}'(P, P_N - P). \tag{6.202}$$

By this approximation,

$$N^{1/2}(\hat{\theta}_N - \theta^*) \approx \mathfrak{F}'(P, N^{1/2}(P_N - P)), \tag{6.203}$$

and we can use $\mathfrak{F}'(P, N^{1/2}(P_N - P))$ to derive asymptotics of $N^{1/2}(\hat{\theta}_N - \theta^*)$.

Suppose, further, that $\mathfrak{F}'(P,\cdot)$ is linear, i.e., $\mathfrak{F}(\cdot)$ is Gâteaux differentiable at P. Then, since $P_N = N^{-1}\sum_{j=1}^N \Delta(Z^j)$, we have that

$$\mathfrak{F}'(P, P_N - P) = \frac{1}{N}\sum_{j=1}^N IF_{\mathfrak{F}}(Z^j), \tag{6.204}$$

where

$$IF_{\mathfrak{F}}(z) := \sum_{j=1}^N \mathfrak{F}'(P, \Delta(z) - P) \tag{6.205}$$

is the so-called *influence function* (also called influence curve) of $\mathfrak{F}$.

It follows from the linearity of $\mathfrak{F}'(P,\cdot)$ that $\mathbb{E}_P[IF_{\mathfrak{F}}(Z)] = 0$. Indeed, linearity of $\mathfrak{F}'(P,\cdot)$ means that it is a linear functional and hence can be represented as

$$\mathfrak{F}'(P, Q-P) = \int g\, d(Q-P) = \int g\, dQ - \mathbb{E}_P[g(Z)]$$

for some function g in an appropriate functional space. Consequently, $IF_{\mathfrak{F}}(z) = g(z) - \mathbb{E}_P[g(Z)]$, and hence

$$\mathbb{E}_P[IF_{\mathfrak{F}}(Z)] = \mathbb{E}_P\{g(Z) - \mathbb{E}_P[g(Z)]\} = 0.$$

Then by the CLT we have that $N^{-1/2}\sum_{j=1}^N IF_{\mathfrak{F}}(Z^j)$ converges in distribution to normal with zero mean and variance $\mathbb{E}_P[IF_{\mathfrak{F}}(Z)^2]$. This suggests the following asymptotics:

$$N^{1/2}(\hat{\theta}_N - \theta^*) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mathbb{E}_P[IF_{\mathfrak{F}}(Z)^2]\right). \tag{6.206}$$

It should be mentioned at this point that the above derivations *do not* prove in a rigorous way validity of the asymptotics (6.206). The main technical difficulty is to give a rigorous justification for the approximation (6.203) leading to the corresponding convergence in distribution. This can be compared with the Delta method, discussed in section 7.2.7 and applied in section 5.1, where first (and second) order approximations were derived in functional spaces rather than spaces of measures. Anyway, formula (6.206) gives correct asymptotics and is routinely used in statistical applications.

Let us consider, for example, the statical functional

$$\mathfrak{F}(P) := \mathbb{E}_P\big[Z - \mathbb{E}_P[Z]\big]_+, \tag{6.207}$$

associated with $\sigma_1^+[Z]$. Denote $\mu := \mathbb{E}_P[Z]$. Then

$$\begin{aligned}\mathfrak{F}(P + t(Q-P)) - \mathfrak{F}(P) = t\left(\mathbb{E}_Q\big[Z-\mu\big]_+ - \mathbb{E}_P\big[Z-\mu\big]_+\right)\\ + \mathbb{E}_P\big[Z - \mu - t(\mathbb{E}_Q[Z]-\mu)\big]_+ + o(t).\end{aligned}$$

Moreover, the right-side derivative at $t = 0$ of the second term in the right-hand side of the above equation is $(1 - H_Z(\mu))(\mathbb{E}_Q[Z] - \mu)$, provided that the cdf $H_Z(z)$ is continuous at $z = \mu$. It follows that if the cdf $H_Z(z)$ is continuous at $z = \mu$, then

$$\mathfrak{F}'(P, Q-P) = \mathbb{E}_Q\big[Z-\mu\big]_+ - \mathbb{E}_P\big[Z-\mu\big]_+ + (1 - H_Z(\mu))(\mathbb{E}_Q[Z] - \mu),$$

and hence

$$IF_{\mathfrak{F}}(z) = [z-\mu]_+ - \mathbb{E}_P\big[Z-\mu\big]_+ + (1 - H_Z(\mu))(z - \mu). \tag{6.208}$$

It can be seen now that $\mathbb{E}_P[IF_{\mathfrak{F}}(Z)] = 0$ and

$$\mathbb{E}_P[IF_{\mathfrak{F}}(Z)^2] = \mathbb{V}\mathrm{ar}\big\{[Z-\mu]_+ + (1 - H_Z(\mu))(Z-\mu)\big\}.$$

That is, the asymptotics (6.206) here are exactly the same as the ones derived in the previous section 6.5.2 (compare with (6.195)).

In a similar way, it is possible to compute the influence function of the statistical functional defined in (6.200), associated with $\sigma_p^+[Z]$, for $p > 1$. For example, for $p = 2$ the corresponding influence function can be computed, provided that the cdf $H_Z(z)$ is continuous at $z = \mu$, as

$$IF_{\mathfrak{F}}(z) = \frac{1}{2\theta^*}\Big([z-\mu]_+^2 - \theta^{*2} + 2\kappa(1 - H_Z(\mu))(z-\mu)\Big), \tag{6.209}$$

where $\theta^* := \mathfrak{F}(P) = (\mathbb{E}_P[Z-\mu]_+^2)^{1/2}$ and $\kappa := \mathbb{E}_P[Z-\mu]_+ = \frac{1}{2}\mathbb{E}_P|Z-\mu|$.

6.6 The Problem of Moments

Due to the duality representation (6.37) of a coherent risk measure, the corresponding risk averse optimization problem (6.128) can be written as the minimax problem (6.131). So far, risk measures were defined on an appropriate functional space, which in turn was dependent on a reference probability distribution. One can take an opposite point of view by defining a min-max problem of the form

$$\operatorname*{Min}_{x\in X} \sup_{P\in\mathfrak{M}} \mathbb{E}_P[f(x,\omega)] \tag{6.210}$$

in a direct way for a specified set $\mathfrak{M}$ of probability measures on a measurable space $(\Omega, \mathcal{F})$. Note that we do not assume in this section existence of a reference measure P and do not work in a functional space of corresponding density functions. In fact, it will be essential here to consider discrete measures on the space $(\Omega, \mathcal{F})$. We denote by $\bar{\mathfrak{P}}$ the set of probability measures[49] on $(\Omega, \mathcal{F})$ and $\mathbb{E}_P[f(x,\omega)]$ is given by the integral

$$\mathbb{E}_P[f(x,\omega)] = \int_\Omega f(x,\omega)dP(\omega).$$

The set $\mathfrak{M}$ can be viewed as an *uncertainty set* for the underlying probability distribution. Of course, there are various ways to define the uncertainty set $\mathfrak{M}$. In some situations, it is reasonable to assume that we have knowledge about certain moments of the corresponding probability distribution. That is, the set $\mathfrak{M}$ is defined by moment constraints as follows:

$$\mathfrak{M} := \left\{ P \in \bar{\mathfrak{P}} : \begin{array}{l} \mathbb{E}_P[\psi_i(\omega)] = b_i,\ i = 1,\dots,p, \\ \mathbb{E}_P[\psi_i(\omega)] \le b_i,\ i = p+1,\dots,q \end{array} \right\}, \tag{6.211}$$

where $\psi_i : \Omega \to \mathbb{R}$, $i = 1,\dots,q$, are measurable functions. Note that the condition $P \in \bar{\mathfrak{P}}$, i.e., that P is a *probability* measure, can be formulated explicitly as the constraint[50] $\int_\Omega dP = 1$, $P \succeq 0$.

We assume that every finite subset of Ω is $\mathcal{F}$-measurable. This is a mild assumption. For example, if Ω is a metric space equipped with its Borel sigma algebra, then this certainly holds true. We denote by $\bar{\mathfrak{P}}^*_m$ the set of probability measures on $(\Omega, \mathcal{F})$ having a finite support of at most m points. That is, every measure $P \in \bar{\mathfrak{P}}^*_m$ can be represented in the form $P = \sum_{i=1}^m \alpha_i \Delta(\omega_i)$, where α_i are nonnegative numbers summing up to one and $\Delta(\omega)$ denotes measure of mass one at the point $\omega \in \Omega$. Similarly, we denote $\mathfrak{M}^*_m := \mathfrak{M} \cap \bar{\mathfrak{P}}^*_m$. Note that the set $\mathfrak{M}$ is convex while, for a fixed m, the set $\mathfrak{M}^*_m$ is not necessarily convex. By Theorem 7.32, to any $P \in \mathfrak{M}$ corresponds a probability measure $Q \in \bar{\mathfrak{P}}$ with a finite support of at most $q+1$ points such that $\mathbb{E}_P[\psi_i(\omega)] = \mathbb{E}_Q[\psi_i(\omega)]$, $i = 1,\dots,q$. That is, if the set $\mathfrak{M}$ is nonempty, then its subset $\mathfrak{M}^*_{q+1}$ is also nonempty. Consider the function

$$g(x) := \sup_{P\in\mathfrak{M}} \mathbb{E}_P[f(x,\omega)]. \tag{6.212}$$

Proposition 6.40. *For any $x \in X$ we have that*

$$g(x) = \sup_{P\in\mathfrak{M}^*_{q+1}} \mathbb{E}_P[f(x,\omega)]. \tag{6.213}$$

[49]The set $\bar{\mathfrak{P}}$ of probability measures should be distinguished from the set $\mathfrak{P}$ of probability density functions used before.

[50]Recall that the notation $P \succeq 0$ means that P is a nonnegative (not necessarily probability) measure on $(\Omega, \mathcal{F})$.

Proof. If the set $\mathfrak{M}$ is empty, then its subset $\mathfrak{M}^*_{q+1}$ is also empty, and hence $g(x)$ as well as the optimal value of the right-hand side of (6.213) are equal to $+\infty$. So suppose that $\mathfrak{M}$ is nonempty. Consider a point $x \in X$ and $P \in \mathfrak{M}$. By Theorem 7.32 there exists $Q \in \mathfrak{M}^*_{q+2}$ such that $\mathbb{E}_P[f(x,\omega)] = \mathbb{E}_Q[f(x,\omega)]$. It follows that $g(x)$ is equal to the maximum of $\mathbb{E}_P[f(x,\omega)]$ over $P \in \mathfrak{M}^*_{q+2}$, which in turn is equal to the optimal value of the problem

$$\begin{aligned} \operatorname*{Max}_{\substack{\omega_1,\dots,\omega_m \in \Omega \\ \alpha \in \mathbb{R}^m_+}} \ & \sum_{j=1}^m \alpha_j f(x,\omega_j) \\ \text{s.t.} \ & \sum_{j=1}^m \alpha_j \psi_i(\omega_j) = b_i, \ i = 1,\dots,p, \\ & \sum_{j=1}^m \alpha_j \psi_i(\omega_j) \le b_i, \ i = p+1,\dots,q, \\ & \sum_{j=1}^m \alpha_j = 1, \end{aligned} \tag{6.214}$$

where $m := q+2$. For fixed $\omega_1,\dots,\omega_m \in \Omega$, the above is a linear programming problem. Its feasible set is bounded and its optimum is attained at an extreme point of its feasible set which has at most $q+1$ nonzero components of α. Therefore it suffices to take the maximum over $P \in \mathfrak{M}^*_{q+1}$. $\square$

For a given $x \in X$, the (Lagrangian) dual of the problem

$$\operatorname*{Max}_{P \in \mathfrak{M}} \mathbb{E}_P[f(x,\omega)] \tag{6.215}$$

is the problem

$$\operatorname*{Min}_{\lambda \in \mathbb{R}\times\mathbb{R}^p\times\mathbb{R}^{q-p}_+} \sup_{P \succeq 0} L_x(P,\lambda), \tag{6.216}$$

where

$$L_x(P,\lambda) := \textstyle\int_\Omega f(x,\omega)dP(\omega) + \lambda_0\big(1 - \int_\Omega dP(\omega)\big) + \sum_{i=1}^q \lambda_i \big(b_i - \int_\Omega \psi_i(\omega)dP(\omega)\big).$$

It is straightforward to verify that

$$\sup_{P \succeq 0} L_x(P,\lambda) = \begin{cases} \lambda_0 + \sum_{i=1}^q b_i\lambda_i & \text{if } f(x,\omega) - \lambda_0 - \sum_{i=1}^q \lambda_i\psi_i(\omega) \le 0, \\ +\infty & \text{otherwise.} \end{cases}$$

The last assertion follows since for any $\bar\omega \in \Omega$ and $\alpha > 0$ we can take $P := \alpha\Delta(\bar\omega)$, in which case

$$\mathbb{E}_P\left[f(x,\omega) - \lambda_0 - \sum_{i=1}^q \lambda_i\psi_i(\omega)\right] = \alpha\left[f(x,\bar\omega) - \lambda_0 - \sum_{i=1}^q \lambda_i\psi_i(\bar\omega)\right].$$

Consequently, we can write the dual problem (6.216) in the form

$$\begin{aligned}&\operatorname*{Min}_{\lambda\in\mathbb{R}\times\mathbb{R}^p\times\mathbb{R}_+^{q-p}} \lambda_0+\sum_{i=1}^q b_i\lambda_i\\&\qquad\text{s.t. } \lambda_0+\sum_{i=1}^q \lambda_i\psi_i(\omega)\geq f(x,\omega),\ \ \omega\in\Omega.\end{aligned}\tag{6.217}$$

If the set Ω is finite, then problem (6.215) and its dual (6.217) are linear programming problems. In that case, there is no duality gap between these problems unless both are infeasible. If the set Ω is infinite, then the dual problem (6.217) becomes a linear semi-infinite programming problem. In that case, one needs to verify some regularity conditions in order to ensure the no-duality-gap property. One such regularity condition will be, "the dual problem (6.217) has a nonempty and bounded set of optimal solutions" (see Theorem 7.8). Another regularity condition ensuring the no-duality-gap property is, "the set Ω is a compact metric space equipped with its Borel sigma algebra and functions $\psi_i(\cdot)$, $i=1,\dots,q$, and $f(x,\cdot)$ are continuous on Ω."

If for every $x\in X$ there is no duality gap between problems (6.215) and (6.217), then the corresponding min-max problem (6.210) is equivalent to the following semi-infinite programming problem:

$$\begin{aligned}&\operatorname*{Min}_{x\in X,\ \lambda\in\mathbb{R}\times\mathbb{R}^p\times\mathbb{R}_+^{q-p}} \lambda_0+\sum_{i=1}^q b_i\lambda_i\\&\qquad\text{s.t. } \lambda_0+\sum_{i=1}^q \lambda_i\psi_i(\omega)\geq f(x,\omega),\ \ \omega\in\Omega.\end{aligned}\tag{6.218}$$

Remark 23. Let Ω be a nonempty measurable subset of $\mathbb{R}^d$, equipped with its Borel sigma algebra, and let $\mathfrak{M}$ be the set of *all* probability measures supported on Ω. Then by the above analysis we have that it suffices in problem (6.210) to take the maximum over measures of mass one, and hence problem (6.210) is equivalent to the following (deterministic) minimax problem:

$$\operatorname*{Min}_{x\in X}\sup_{\omega\in\Omega} f(x,\omega).\tag{6.219}$$

6.7 Multistage Risk Averse Optimization

In this section we discuss an extension of risk averse optimization to a multistage setting. In order to simplify the presentation we start our analysis with a discrete process in which evolution of the state of the system is represented by a scenario tree.

6.7.1 Scenario Tree Formulation

Consider a scenario tree representation of evolution of the corresponding data process (see section 3.1.3). The basic idea of multistage stochastic programming is that if we are currently at a state of the system at stage t, represented by a node of the scenario tree, then our decision at that node is based on our knowledge about the next possible realizations of the data process, which are represented by its children nodes at stage $t + 1$. In the risk neutral approach we optimize the corresponding conditional expectation of the objective function. This allows us to write the associated dynamic programming equations. This idea can be extended to optimization of a risk measure conditional on a current state of the system. We now discuss such construction in detail.

As in section 3.1.3, we denote by Ω_t the set of all nodes at stage $t = 1, \dots, T$, by $K_t := |\Omega_t|$ the cardinality of Ω_t and by C_a the set of children nodes of a node a of the tree. Note that $\{C_a\}_{a\in\Omega_t}$ forms a partition of the set Ω_{t+1}, i.e., $C_a \cap C_{a'} = \emptyset$ if $a \neq a'$ and $\Omega_{t+1} = \cup_{a\in\Omega_t} C_a$, $t = 1, \dots, T-1$. With the set Ω_T we associate sigma algebra $\mathcal{F}_T$ of *all* its subsets. Let $\mathcal{F}_{T-1}$ be the subalgebra of $\mathcal{F}_T$ generated by sets C_a, $a \in \Omega_{T-1}$, i.e., these sets form the set of elementary events of $\mathcal{F}_{T-1}$. (Recall that $\{C_a\}_{a\in\Omega_{T-1}}$ forms a partition of Ω_T.) By this construction, there is a one-to-one correspondence between elementary events of $\mathcal{F}_{T-1}$ and the set Ω_{T-1} of nodes at stage $T-1$. By continuing this process we construct a sequence of sigma algebras $\mathcal{F}_1 \subset \cdots \subset \mathcal{F}_T$. (Such a sequence of nested sigma algebras is called *filtration*.) Note that $\mathcal{F}_1$ corresponds to the unique root node and hence $\mathcal{F}_1 = \{\emptyset, \Omega_T\}$. In this construction, there is a one-to-one correspondence between nodes of Ω_t and elementary events of the sigma algebra $\mathcal{F}_t$, and hence we can identify every node $a \in \Omega_t$ with an elementary event of $\mathcal{F}_t$. By taking all children of every node of C_a at later stages, we eventually can identify with C_a a subset of Ω_T.

Suppose, further, that there is a probability distribution defined on the scenario tree. As discussed in section 3.1.3, such probability distribution can be defined by introducing conditional probabilities of going from a node of the tree to its children nodes. That is, with a node $a \in \Omega_t$ is associated a probability vector[51] $p^a \in \mathbb{R}^{|C_a|}$ of conditional probabilities of moving from a to nodes of C_a. Equipped with probability vector p^a, the set C_a becomes a probability space, with the corresponding sigma algebra of all subsets of C_a, and any function $Z : C_a \to \mathbb{R}$ can be viewed as a random variable. Since the space of functions $Z : C_a \to \mathbb{R}$ can be identified with the space $\mathbb{R}^{|C_a|}$, we identify such random variable Z with an element of the vector space $\mathbb{R}^{|C_a|}$. With every $Z \in \mathbb{R}^{|C_a|}$ is associated the expectation $\mathbb{E}_{p^a}[Z]$, which can be considered as a conditional expectation given that we are currently at node a.

Now with every node a at stage $t = 1, \dots, T-1$ we associate a risk measure $\rho^a(Z)$ defined on the space of functions $Z : C_a \to \mathbb{R}$, that is, we choose a family of risk measures

$$\rho^a : \mathbb{R}^{|C_a|} \to \mathbb{R}, \quad a \in \Omega_t, \ t = 1, \dots, T-1. \tag{6.220}$$

Of course, there are many ways to define such risk measures. For instance, for a given probability distribution on the scenario tree, we can use conditional expectations

$$\rho^a(Z) := \mathbb{E}_{p^a}[Z], \quad a \in \Omega_t, \ t = 1, \dots, T-1. \tag{6.221}$$

Such choice of risk measures ρ^a leads to the risk neutral formulation of a corresponding multistage stochastic program. For a risk averse approach we can use any class of coherent risk measures discussed in section 6.3.2, as, for example,

$$\rho^a[Z] := \inf_{t\in\mathbb{R}} \left\{ t + \lambda_a^{-1} \mathbb{E}_{p^a}\big[Z - t\big]_+ \right\}, \quad \lambda_a \in (0, 1), \tag{6.222}$$

[51] A vector $p = (p_1, \dots, p_n) \in \mathbb{R}^n$ is said to be a *probability vector* if all its components p_i are nonnegative and $\sum_{i=1}^n p_i = 1$. If $Z = (Z_1, \dots, Z_n) \in \mathbb{R}^n$ is viewed as a random variable, then its expectation with respect to p is $\mathbb{E}_p[Z] = \sum_{i=1}^n p_i Z_i$.

corresponding to AV@R risk measure and

$$\rho^a[Z] := \mathbb{E}_{p^a}[Z] + c_a \mathbb{E}_{p^a}\big[Z - \mathbb{E}_{p^a}[Z]\big]_+, \quad c_a \in [0,1], \tag{6.223}$$

corresponding to the absolute semideviation risk measure.

Since Ω_{t+1} is the union of the disjoint sets C_a, $a \in \Omega_t$, we can write $\mathbb{R}^{K_{t+1}}$ as the Cartesian product of the spaces $\mathbb{R}^{|C_a|}$, $a \in \Omega_t$. That is, $\mathbb{R}^{K_{t+1}} = \mathbb{R}^{|C_{a_1}|} \times \cdots \times \mathbb{R}^{|C_{a_{K_t}}|}$, where $\{a_1, \dots, a_{K_t}\} = \Omega_t$. Define the mappings

$$\rho_{t+1} := (\rho^{a_1}, \dots, \rho^{a_{K_t}}) : \mathbb{R}^{K_{t+1}} \to \mathbb{R}^{K_t}, \quad t = 1, \dots, T-1, \tag{6.224}$$

associated with risk measures ρ^a. Recall that the set Ω_{t+1} of nodes at stage $t+1$ is identified with the set of elementary events of sigma algebra $\mathcal{F}_{t+1}$, and its sigma subalgebra $\mathcal{F}_t$ is generated by sets C_a, $a \in \Omega_t$.

We denote by $\mathcal{Z}_T$ the space of all functions $Z : \Omega_T \to \mathbb{R}$. As mentioned, we can identify every such function with a vector of the space $\mathbb{R}^{K_T}$, i.e., the space $\mathcal{Z}_T$ can be identified with the space $\mathbb{R}^{K_T}$. We have that a function $Z : \Omega_T \to \mathbb{R}$ is $\mathcal{F}_{T-1}$-measurable iff it is constant on every set C_a, $a \in \Omega_{T-1}$. We denote by $\mathcal{Z}_{T-1}$ the subspace of $\mathcal{Z}_T$ formed by $\mathcal{F}_{T-1}$-measurable functions. The space $\mathcal{Z}_{T-1}$ can be identified with $\mathbb{R}^{K_{T-1}}$. And so on, we can construct a sequence $\mathcal{Z}_t$, $t = 1, \dots, T$, of spaces of $\mathcal{F}_t$-measurable functions $Z : \Omega_T \to \mathbb{R}$ such that $\mathcal{Z}_1 \subset \cdots \subset \mathcal{Z}_T$ and each $\mathcal{Z}_t$ can be identified with the space $\mathbb{R}^{K_t}$. Recall that $K_1 = 1$, and hence $\mathcal{Z}_1$ can be identified with $\mathbb{R}$. We view the mapping ρ_{t+1}, defined in (6.224), as a mapping from the space $\mathcal{Z}_{t+1}$ into the space $\mathcal{Z}_t$. Conversely, with any mapping $\rho_{t+1} : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$ we can associate a family of risk measures of the form (6.220).

We say that a mapping $\rho_{t+1} : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$ is a *conditional risk mapping* if it satisfies the following conditions:[52]

(R′1) Convexity:

$$\rho_{t+1}(\alpha Z + (1-\alpha)Z') \preceq \alpha \rho_{t+1}(Z) + (1-\alpha)\rho_{t+1}(Z')$$

for any $Z, Z' \in \mathcal{Z}_{t+1}$ and $\alpha \in [0,1]$.

(R′2) Monotonicity: If $Z, Z' \in \mathcal{Z}_{t+1}$ and $Z \succeq Z'$, then $\rho_{t+1}(Z) \succeq \rho_{t+1}(Z')$.

(R′3) Translation equivariance: If $Y \in \mathcal{Z}_t$ and $Z \in \mathcal{Z}_{t+1}$, then $\rho_{t+1}(Z+Y) = \rho_{t+1}(Z) + Y$.

(R′4) Positive homogeneity: If $\alpha \geq 0$ and $Z \in \mathcal{Z}_{t+1}$, then $\rho_{t+1}(\alpha Z) = \alpha \rho_{t+1}(Z)$.

It is straightforward to see that conditions (R′1), (R′2), and (R′4) hold iff the corresponding conditions (R1), (R2), and (R4), defined in section 6.3, hold for every risk measure ρ^a associated with ρ_{t+1}. Also by construction of ρ_{t+1}, we have that condition (R′3) holds iff condition (R3) holds for all ρ^a. That is, ρ_{t+1} is a *conditional risk mapping* iff every corresponding risk measure ρ^a is a *coherent risk measure.*

By Theorem 6.4 with each coherent risk measure ρ^a, $a \in \Omega_t$, is associated a set $\mathfrak{A}(a)$ of probability measures (vectors) such that

$$\rho^a(Z) = \max_{p \in \mathfrak{A}(a)} \mathbb{E}_p[Z]. \tag{6.225}$$

[52]For $Z_1, Z_2 \in \mathcal{Z}_t$ the inequality $Z_2 \succeq Z_1$ is understood componentwise.

Here $Z \in \mathbb{R}^{K_{t+1}}$ is a vector corresponding to function $Z : \Omega_{t+1} \to \mathbb{R}$, and $\mathfrak{A}(a) = \mathfrak{A}_{t+1}(a)$ is a closed convex set of probability vectors $p \in \mathbb{R}^{K_{t+1}}$ such that $p_k = 0$ if $k \in \Omega_{t+1} \setminus C_a$, i.e., all probability measures of $\mathfrak{A}_{t+1}(a)$ are supported on the set C_a. We can now represent the corresponding conditional risk mapping ρ_{t+1} as a maximum of conditional expectations as follows. Let $\nu = (\nu_a)_{a\in\Omega_t}$ be a probability distribution on Ω_t, assigning *positive* probability ν_a to every $a \in \Omega_t$, and define

$$\mathfrak{C}_{t+1} := \left\{ \mu = \sum_{a\in\Omega_t} \nu_a p^a : p^a \in \mathfrak{A}_{t+1}(a) \right\}. \tag{6.226}$$

It is not difficult to see that $\mathfrak{C}_{t+1} \subset \mathbb{R}^{K_{t+1}}$ is a convex set of probability vectors. Moreover, since each $\mathfrak{A}_{t+1}(a)$ is compact, the set $\mathfrak{C}_{t+1}$ is also compact and hence is closed. Consider a probability distribution (measure) $\mu = \sum_{a\in\Omega_t} \nu_a p^a \in \mathfrak{C}_{t+1}$. We have that for $a \in \Omega_t$, the corresponding conditional distribution given the event C_a is p^a, and[53]

$$\mathbb{E}_\mu \left[Z|\mathcal{F}_t \right](a) = \mathbb{E}_{p^a}[Z], \quad Z \in \mathcal{Z}_{t+1}. \tag{6.227}$$

It follows then by (6.225) that

$$\rho_{t+1}(Z) = \max_{\mu\in\mathfrak{C}_{t+1}} \mathbb{E}_\mu \left[Z|\mathcal{F}_t \right], \tag{6.228}$$

where the maximum on the right-hand side of (6.228) is taken pointwise in $a \in \Omega_t$. That is, formula (6.228) means that

$$[\rho_{t+1}(Z)](a) = \max_{p\in\mathfrak{A}_{t+1}(a)} \mathbb{E}_p[Z], \quad Z \in \mathcal{Z}_{t+1},\ a \in \Omega_t. \tag{6.229}$$

Note that in this construction, choice of the distribution ν is arbitrary and any distribution of $\mathfrak{C}_{t+1}$ agrees with the distribution ν on Ω_t.

We are ready now to give a formulation of risk averse multistage programs. For a sequence $\rho_{t+1} : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$, $t = 1, \dots, T-1$, of conditional risk mappings, consider the following risk averse formulation analogous to the nested risk neutral formulation (3.1):

$$\begin{aligned} \operatorname*{Min}_{x_1\in\mathcal{X}_1} f_1(x_1) + \rho_2\Bigg[& \inf_{x_2\in\mathcal{X}_2(x_1,\omega)} f_2(x_2,\omega) + \cdots \\ & + \rho_{T-1}\Big[\inf_{x_{T-1}\in\mathcal{X}_T(x_{T-2},\omega)} f_{T-1}(x_{T-1},\omega) \\ & + \rho_T\Big[\inf_{x_T\in\mathcal{X}_T(x_{T-1},\omega)} f_T(x_T,\omega) \Big]\Big]\Bigg]. \end{aligned} \tag{6.230}$$

Here ω is an element of $\Omega := \Omega_T$, the objective functions $f_t : \mathbb{R}^{n_{t-1}} \times \Omega \to \mathbb{R}$ are real valued functions, and $\mathcal{X}_t : \mathbb{R}^{n_{t-1}} \times \Omega \rightrightarrows \mathbb{R}^{n_t}$, $t = 2, \dots, T$, are multifunctions such that $f_t(x_t, \cdot)$ and $\mathcal{X}_t(x_{t-1}, \cdot)$ are $\mathcal{F}_t$-measurable for all x_t and x_{t-1}. Note that if the corresponding risk measures ρ^a are defined as conditional expectations (6.221), then the multistage problem (6.230) coincides with the risk neutral multistage problem (3.1).

[53]Recall that the conditional expectation $\mathbb{E}_\mu[\,\cdot\,|\mathcal{F}_t]$ is a mapping from $\mathcal{Z}_{t+1}$ into $\mathcal{Z}_t$.

There are several ways in which the nested formulation (6.230) can be formalized. Similarly to (3.3), we can write problem (6.230) in the form

$$\begin{array}{ll} \operatorname*{Min}_{x_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_T} & f_1(x_1) + \rho_2\Big[f_2(\boldsymbol{x}_2(\omega),\omega) + \cdots \\ & \qquad + \rho_{T-1}\big[f_{T-1}(\boldsymbol{x}_{T-1}(\omega),\omega) + \rho_T[f_T(\boldsymbol{x}_T(\omega),\omega)]\big]\Big] \\ \text{s.t.} & x_1 \in \mathcal{X}_1,\ \boldsymbol{x}_t(\omega) \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\omega),\omega),\ t = 2,\dots,T. \end{array} \tag{6.231}$$

Optimization in (6.231) is performed over functions $\boldsymbol{x}_t : \Omega \to \mathbb{R}$, $t = 1,\dots,T$, satisfying the corresponding constraints, which imply that each $\boldsymbol{x}_t(\omega)$ is $\mathcal{F}_t$-measurable and hence each $f_t(\boldsymbol{x}_t(\omega),\omega)$ is $\mathcal{F}_t$-measurable. The requirement for $\boldsymbol{x}_t(\omega)$ to be $\mathcal{F}_t$-measurable is another way of formulating the *nonanticipativity constraints*. Therefore, it can be viewed that the optimization in (6.231) is performed over feasible policies.

Consider the function $\varrho : \mathcal{Z}_1 \times \cdots \times \mathcal{Z}_T \to \mathbb{R}$ defined as

$$\varrho(Z_1,\dots,Z_T) := Z_1 + \rho_2\Big[Z_2 + \cdots + \rho_{T-1}\big[Z_{T-1} + \rho_T[Z_T]\big]\Big]. \tag{6.232}$$

By condition (R$'$3) we have that

$$\rho_{T-1}\big[Z_{T-1} + \rho_T[Z_T]\big] = \rho_{T-1} \circ \rho_T\big[Z_{T-1} + Z_T\big].$$

By continuing this process we obtain that

$$\varrho(Z_1,\dots,Z_T) = \bar{\rho}(Z_1 + \ldots + Z_T), \tag{6.233}$$

where $\bar{\rho} := \rho_2 \circ \cdots \circ \rho_T$. We refer to $\bar{\rho}$ as the *composite risk measure*. That is,

$$\bar{\rho}(Z_1 + \cdots + Z_T) = Z_1 + \rho_2\Big[Z_2 + \cdots + \rho_{T-1}\big[Z_{T-1} + \rho_T[Z_T]\big]\Big], \tag{6.234}$$

defined for $Z_t \in \mathcal{Z}_t$, $t = 1,\dots,T$. Recall that $\mathcal{Z}_1$ is identified with $\mathbb{R}$, and hence Z_1 is a real number and $\bar{\rho} : \mathcal{Z}_T \to \mathbb{R}$ is a real valued function. Conditions (R$'$1)–(R$'$4) imply that $\bar{\rho}$ is a coherent risk measure.

As above, we have that since $f_{T-1}(\boldsymbol{x}_{T-1}(\omega),\omega)$ is $\mathcal{F}_{T-1}$-measurable, it follows by condition (R$'$3) that

$$f_{T-1}(\boldsymbol{x}_{T-1}(\omega),\omega) + \rho_T\left[f_T(\boldsymbol{x}_T(\omega),\omega)\right] = \rho_T\left[f_{T-1}(\boldsymbol{x}_{T-1}(\omega),\omega) + f_T(\boldsymbol{x}_T(\omega),\omega)\right].$$

Continuing this process backward, we obtain that the objective function of (6.231) can be formulated using the composite risk measure. That is, problem (6.231) can be written in the form

$$\begin{array}{ll} \operatorname*{Min}_{x_1,\boldsymbol{x}_2,\dots,\boldsymbol{x}_T} & \bar{\rho}\big[f_1(x_1) + f_2(\boldsymbol{x}_2(\omega),\omega) + \cdots + f_T(\boldsymbol{x}_T(\omega),\omega)\big] \\ \text{s.t.} & x_1 \in \mathcal{X}_1,\ \boldsymbol{x}_t(\omega) \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\omega),\omega),\ t = 2,\dots,T. \end{array} \tag{6.235}$$

If the conditional risk mappings are defined as the respective conditional expectations, then the composite risk measure $\bar{\rho}$ becomes the corresponding expectation operator, and (6.235) coincides with the multistage program written in the form (3.3). Unfortunately, it is not easy to write the composite risk measure $\bar{\rho}$ in a closed form even for relatively simple conditional risk mappings other than conditional expectations.

An alternative approach to formalize the nested formulation (6.230) is to write *dynamic programming equations*. That is, for the last period T we have

$$Q_T(x_{T-1},\omega) := \inf_{x_T \in \mathcal{X}_T(x_{T-1},\omega)} f_T(x_T,\omega), \tag{6.236}$$

$$\mathcal{Q}_T(x_{T-1},\omega) := \rho_T[Q_T(x_{T-1},\omega)], \tag{6.237}$$

and for $t = T-1,\dots,2$, we recursively apply the conditional risk measures

$$\mathcal{Q}_t(x_{t-1},\omega) := \rho_t\left[Q_t(x_{t-1},\omega)\right], \tag{6.238}$$

where

$$Q_t(x_{t-1},\omega) := \inf_{x_t \in \mathcal{X}_t(x_{t-1},\omega)} \Big\{ f_t(x_t,\omega) + \mathcal{Q}_{t+1}(x_t,\omega) \Big\}. \tag{6.239}$$

Of course, equations (6.238) and (6.239) can be combined into one equation:[54]

$$Q_t(x_{t-1},\omega) = \inf_{x_t \in \mathcal{X}_t(x_{t-1},\omega)} \Big\{ f_t(x_t,\omega) + \rho_{t+1}\left[Q_{t+1}(x_t,\omega)\right] \Big\}. \tag{6.240}$$

Finally, at the first stage we solve the problem

$$\operatorname*{Min}_{x_1 \in \mathcal{X}_1} f_1(x_1) + \rho_2[Q_2(x_1,\omega)]. \tag{6.241}$$

It is important to emphasize that conditional risk mappings $\rho_t(Z)$ are defined on real valued functions $Z(\omega)$. Therefore, it is implicitly assumed in the above equations that the cost-to-go (value) functions $Q_t(x_{t-1},\omega)$ are real valued. In particular, this implies that the considered problem should have relatively complete recourse. Also, in the above development of the dynamic programming equations, the monotonicity condition (R$'$2) plays a crucial role, because only then we can move the optimization under the risk operation.

Remark 24. By using representation (6.228), we can write the dynamic programming equations (6.240) in the form

$$Q_t(x_{t-1},\omega) = \inf_{x_t \in \mathcal{X}_t(x_{t-1},\omega)} \left\{ f_t(x_t,\omega) + \sup_{\mu \in \mathfrak{C}_{t+1}} \mathbb{E}_\mu\left[Q_{t+1}(x_t)\middle|\mathcal{F}_t\right](\omega) \right\}. \tag{6.242}$$

Note that the left- and right-hand-side functions in (6.242) are $\mathcal{F}_t$-measurable, and hence this equation can be written in terms of $a \in \Omega_t$ instead of $\omega \in \Omega$. Recall that every $\mu \in \mathfrak{C}_{t+1}$ is representable in the form $\mu = \sum_{a \in \Omega_t} \nu_a p^a$ (see (6.226)) and that

$$\mathbb{E}_\mu\left[Q_{t+1}(x_t)\middle|\mathcal{F}_t\right](a) = \mathbb{E}_{p^a}[Q_{t+1}(x_t)], \quad a \in \Omega_t. \tag{6.243}$$

We say that the problem is *convex* if the functions $f_t(\cdot,\omega)$, $Q_t(\cdot,\omega)$ and the sets $\mathcal{X}_t(x_{t-1},\omega)$ are convex for every $\omega \in \Omega$ and $t = 1,\dots,T$. If the problem is convex, then (since the

[54]With some abuse of the notation we write $Q_{t+1}(x_t,\omega)$ for the value of $Q_{t+1}(x_t)$ at $\omega \in \Omega$, and $\rho_{t+1}\left[Q_{t+1}(x_t,\omega)\right]$ for $\rho_{t+1}\left[Q_{t+1}(x_t)\right](\omega)$.

set $\mathfrak{C}_{t+1}$ is convex compact) the inf and sup operators on the right-hand side of (6.242) can be interchanged to obtain a dual problem, and for a given x_{t-1} and every $a \in \Omega_t$ the dual problem has an optimal solution $\bar{p}^a \in \mathfrak{A}_{t+1}(a)$. Consequently, for $\bar{\mu}_{t+1} := \sum_{a\in\Omega_t} \nu_a \bar{p}^a$ an optimal solution of the original problem and the corresponding cost-to-go functions satisfy the following dynamic programming equations:

$$Q_t(x_{t-1}, \omega) = \inf_{x_t \in \mathcal{X}_t(x_{t-1},\omega)} \Big\{ f_t(x_t, \omega) + \mathbb{E}_{\bar{\mu}_{t+1}}\big[Q_{t+1}(x_t) | \mathcal{F}_t \big](\omega) \Big\}. \tag{6.244}$$

Moreover, it is possible to choose the "worst case" distributions $\bar{\mu}_{t+1}$ in a consistent way, i.e., such that each $\bar{\mu}_{t+1}$ coincides with $\bar{\mu}_t$ on $\mathcal{F}_t$. That is, consider the first-stage problem (6.241). We have that (recall that at the first stage there is only one node, $\mathcal{F}_1 = \{\emptyset, \Omega\}$ and $\mathfrak{C}_2 = \mathfrak{A}_2$)

$$\rho_2[Q_2(x_1)] = \sup_{\mu\in\mathfrak{C}_2} \mathbb{E}_\mu[Q_2(x_1)|\mathcal{F}_1] = \sup_{\mu\in\mathfrak{C}_2} \mathbb{E}_\mu[Q_2(x_1)]. \tag{6.245}$$

By convexity and since $\mathfrak{C}_2$ is compact, we have that there is $\bar{\mu}_2 \in \mathfrak{C}_2$(an optimal solution of the dual problem) such that the optimal value of the first-stage problem is equal to the optimal value and the set of optimal solutions of the first-stage problem is contained in the set of optimal solutions of the problem

$$\operatorname*{Min}_{x_1\in\mathcal{X}_1} \mathbb{E}_{\bar{\mu}_2}[Q_2(x_1)]. \tag{6.246}$$

Let $\bar{x}_1$ be an optimal solution of the first-stage problem. Then we can choose $\bar{\mu}_3 \in \mathfrak{C}_3$, of the form $\bar{\mu}_3 := \sum_{a\in\Omega_2} \nu_a \bar{p}^a$ such that (6.244) holds with $t = 2$ and $x_1 = \bar{x}_1$. Moreover, we can take the probability measure $\nu = (\nu_a)_{a\in\Omega_2}$ to be the same as $\bar{\mu}_2$, and hence to ensure that $\bar{\mu}_3$ coincides with $\bar{\mu}_2$ on $\mathcal{F}_2$. Next, for every node $a \in \Omega_2$ choose a corresponding (second-stage) optimal solution and repeat the construction to produce an appropriate $\bar{\mu}_4 \in \mathfrak{C}_4$, and so on for later stages.

In that way, assuming existence of optimal solutions, we can construct a probability distribution $\bar{\mu}_2, \dots, \bar{\mu}_T$ on the considered scenario tree such that the obtained multistage problem, of the standard form (3.1), has the same cost-to-go (value) functions as the original problem (6.230) and has an optimal solution which also is an optimal solution of the problem (6.230). (In that sense, the obtained multistage problem, driven by dynamic programming equations (6.244), is almost equivalent to the original problem.)

Remark 25. Let us define, for every node $a \in \Omega_t$, $t = 1, \dots, T-1$, the corresponding set $\mathfrak{A}(a) = \mathfrak{A}_{t+1}(a)$ to be the set of *all* probability measures (vectors) on the set C_a. (Recall that $C_a \subset \Omega_{t+1}$ is the set of children nodes of a and that all probability measures of $\mathfrak{A}_{t+1}(a)$ are supported on C_a.) Then the maximum on the right-hand side of (6.225) is attained at a measure of mass one at a point of the set C_a. Consequently, by (6.243), for such choice of the sets $\mathfrak{A}_{t+1}(a)$ the dynamic programming equations (6.242) can be written as

$$Q_t(x_{t-1}, a) = \inf_{x_t\in\mathcal{X}_t(x_{t-1},a)} \Big\{ f_t(x_t, a) + \max_{\omega\in C_a} Q_{t+1}(x_t, \omega) \Big\}, \quad a \in \Omega_t. \tag{6.247}$$

It is interesting to note (see Remark 24, page 313) that if the problem is convex, then it is possible to construct a probability distribution (on the considered scenario tree), defined by a sequence $\bar{\mu}_t$, $t = 2, \dots, T$, of consistent probability distributions, such that the obtained (risk neutral) multistage program is almost equivalent to the min-max formulation (6.247).

6.7.2 Conditional Risk Mappings

In this section we discuss a general concept of conditional risk mappings which can be applied to a risk averse formulation of multistage programs. The material of this section can be considered as an extension to an infinite dimensional setting of the developments presented in the previous section. Similarly to the presentation of coherent risk measures, given in section 6.3, we use the framework of $\mathcal{L}_p$ spaces, $p \in [1, +\infty)$. That is, let Ω be a sample space equipped with sigma algebras $\mathcal{F}_1 \subset \mathcal{F}_2$ (i.e., $\mathcal{F}_1$ is subalgebra of $\mathcal{F}_2$) and a probability measure P on $(\Omega, \mathcal{F}_2)$. Consider the spaces $\mathcal{Z}_1 := \mathcal{L}_p(\Omega, \mathcal{F}_1, P)$ and $\mathcal{Z}_2 := \mathcal{L}_p(\Omega, \mathcal{F}_2, P)$. Since $\mathcal{F}_1$ is a subalgebra of $\mathcal{F}_2$, it follows that $\mathcal{Z}_1 \subset \mathcal{Z}_2$.

We say that a mapping $\rho : \mathcal{Z}_2 \to \mathcal{Z}_1$ is a *conditional risk mapping* if it satisfies the following conditions:

(R′1) Convexity:
$$\rho(\alpha Z + (1-\alpha)Z') \preceq \alpha\rho(Z) + (1-\alpha)\rho(Z')$$
for any $Z, Z' \in \mathcal{Z}_2$ and $\alpha \in [0, 1]$.

(R′2) Monotonicity: If $Z, Z' \in \mathcal{Z}_2$ and $Z \succeq Z'$, then $\rho(Z) \succeq \rho(Z')$.

(R′3) Translation equivariance: If $Y \in \mathcal{Z}_1$ and $Z \in \mathcal{Z}_2$, then $\rho(Z + Y) = \rho(Z) + Y$.

(R′4) Positive homogeneity: If $\alpha \geq 0$ and $Z \in \mathcal{Z}_2$, then $\rho(\alpha Z) = \alpha\rho(Z)$.

The above conditions coincide with the respective conditions of the previous section which were defined in a finite dimensional setting. If the sigma algebra $\mathcal{F}_1$ is trivial, i.e., $\mathcal{F}_1 = \{\emptyset, \Omega\}$, then the space $\mathcal{Z}_1$ can be identified with $\mathbb{R}$, and conditions (R′1)–(R′4) define a coherent risk measure. Examples of coherent risk measures, discussed in section 6.3.2, have conditional risk mapping analogues which are obtained by replacing the expectation operator with the corresponding conditional expectation $\mathbb{E}[\,\cdot\,|\mathcal{F}_1]$ operator. Let us look at some examples.

Conditional Expectation. In itself, the conditional expectation mapping $\mathbb{E}[\,\cdot\,|\mathcal{F}_1] : \mathcal{Z}_2 \to \mathcal{Z}_1$ is a conditional risk mapping. Indeed, for any $p \geq 1$ and $Z \in \mathcal{L}_p(\Omega, \mathcal{F}_2, P)$ we have by Jensen inequality that $\mathbb{E}\big[|Z|^p|\mathcal{F}_1\big] \succeq \big|\mathbb{E}[Z|\mathcal{F}_1]\big|^p$, and hence

$$\int_\Omega \big|\mathbb{E}[Z|\mathcal{F}_1]\big|^p dP \leq \int_\Omega \mathbb{E}\big[|Z|^p|\mathcal{F}_1\big] dP = \mathbb{E}\big[|Z|^p\big] < +\infty. \tag{6.248}$$

This shows that, indeed, $\mathbb{E}[\,\cdot\,|\mathcal{F}_1]$ maps $\mathcal{Z}_2$ into $\mathcal{Z}_1$. The conditional expectation is a linear operator, and hence conditions (R′1) and (R′4) follow. The monotonicity condition (R′2) also clearly holds, and condition (R′3) is a property of conditional expectation.

Conditional AV@R. An analogue of the AV@R risk measure can be defined as follows. Let $\mathcal{Z}_i := \mathcal{L}_1(\Omega, \mathcal{F}_i, P)$, $i = 1, 2$. For $\alpha \in (0, 1)$ define mapping $\mathsf{AV@R}_\alpha(\,\cdot\,|\mathcal{F}_1) : \mathcal{Z}_2 \to \mathcal{Z}_1$ as

$$[\mathsf{AV@R}_\alpha(Z|\mathcal{F}_1)](\omega) := \inf_{Y \in \mathcal{Z}_1} \Big\{ Y(\omega) + \alpha^{-1}\mathbb{E}\big[[Z - Y]_+\big|\mathcal{F}_1\big](\omega) \Big\}, \quad \omega \in \Omega. \tag{6.249}$$

It is not difficult to verify that, indeed, this mapping satisfies conditions (R′1)–(R′4). Similarly to (6.68), for $\beta \in [0,1]$ and $\alpha \in (0,1)$, we can also consider the following conditional risk mapping:

$$\rho_{\alpha,\beta|\mathcal{F}_1}(Z) := (1-\beta)\mathbb{E}[Z|\mathcal{F}_1] + \beta \mathsf{AV@R}_\alpha(Z|\mathcal{F}_1). \tag{6.250}$$

Of course, the above conditional risk mapping $\rho_{\alpha,\beta|\mathcal{F}_1}$ corresponds to the coherent risk measure $\rho_{\alpha,\beta}(Z) := (1-\beta)\mathbb{E}[Z] + \beta \mathsf{AV@R}_\alpha(Z)$.

Conditional Mean-Upper-Semideviation. An analogue of the mean-upper-semideviation risk measure (of order p) can be constructed as follows. Let $\mathcal{Z}_i := \mathcal{L}_p(\Omega, \mathcal{F}_i, P)$, $i = 1, 2$. For $c \in [0,1]$ define

$$\rho_{c|\mathcal{F}_1}(Z) := \mathbb{E}[Z|\mathcal{F}_1] + c\left(\mathbb{E}\Big[\big[Z - \mathbb{E}[Z|\mathcal{F}_1]\big]_+^p \big| \mathcal{F}_1\Big]\right)^{1/p}. \tag{6.251}$$

In particular, for $p = 1$ this gives an analogue of the absolute semideviation risk measure.

In the discrete case of scenario tree formulation (discussed in the previous section) the above examples correspond to taking the same respective risk measure at every node of the considered tree at stage $t = 1, \dots, T$.

Consider a conditional risk mapping $\rho : \mathcal{Z}_2 \to \mathcal{Z}_1$. With a set $A \in \mathcal{F}_1$, such that $P(A) \neq 0$, we associate the function

$$\rho_A(Z) := \mathbb{E}[\rho(Z)|A], \quad Z \in \mathcal{Z}_2, \tag{6.252}$$

where $\mathbb{E}[Y|A] := \frac{1}{P(A)}\int_A Y dP$ denotes the conditional expectation of random variable $Y \in \mathcal{Z}_1$ given event $A \in \mathcal{F}_1$. Clearly conditions (R′1)–(R′4) imply that the corresponding conditions (R1)–(R4) hold for ρ_A, and hence ρ_A is a coherent risk measure defined on the space $\mathcal{Z}_2 = \mathcal{L}_p(\Omega, \mathcal{F}_2, P)$. Moreover, for any $B \in \mathcal{F}_1$ we have by (R′3) that

$$\rho_A(Z + \alpha \mathbf{1}_B) := \mathbb{E}[\rho(Z) + \alpha \mathbf{1}_B | A] = \rho_A(Z) + \alpha P(B|A) \quad \forall \alpha \in \mathbb{R}, \tag{6.253}$$

where $P(B|A) = P(B \cap A)/P(A)$.

Since ρ_A is a coherent risk measure, by Theorem 6.4 it can be represented in the form

$$\rho_A(Z) = \sup_{\zeta \in \mathfrak{A}(A)} \int_\Omega \zeta(\omega) Z(\omega) dP(\omega) \tag{6.254}$$

for some set $\mathfrak{A}(A) \subset \mathcal{L}_q(\Omega, \mathcal{F}_2, P)$ of probability density functions. Let us make the following observation:

- Each density $\zeta \in \mathfrak{A}(A)$ is supported on the set A.

Indeed, for any $B \in \mathcal{F}_1$, such that $P(B \cap A) = 0$, and any $\alpha \in \mathbb{R}$, we have by (6.253) that $\rho_A(Z + \alpha \mathbf{1}_B) = \rho_A(Z)$. On the other hand, if there exists $\zeta \in \mathfrak{A}(A)$ such that $\int_B \zeta dP > 0$, then it follows from (6.254) that $\rho_A(Z + \alpha \mathbf{1}_B)$ tends to $+\infty$ as $\alpha \to +\infty$.

Similarly to (6.228), we show now that a conditional risk mapping can be represented as a maximum of a family of conditional expectations. We consider a situation where the subalgebra $\mathcal{F}_1$ has a countable number of elementary events. That is, there is a (countable) partition $\{A_i\}_{i\in\mathbb{N}}$ of the sample space Ω which generates $\mathcal{F}_1$, i.e., $\cup_{i\in\mathbb{N}} A_i = \Omega$, the sets A_i, $i \in \mathbb{N}$, are disjoint and form the family of elementary events of sigma algebra $\mathcal{F}_1$. Since $\mathcal{F}_1$ is a subalgebra of $\mathcal{F}_2$, we have of course that $A_i \in \mathcal{F}_2$, $i \in \mathbb{N}$. We also have that a function $Z : \Omega \to \mathbb{R}$ is $\mathcal{F}_1$-measurable iff it is constant on every set A_i, $i \in \mathbb{N}$.

Consider a conditional risk mapping $\rho : \mathcal{Z}_2 \to \mathcal{Z}_1$. Let

$$\mathfrak{N} := \{i \in \mathbb{N} : P(A_i) \neq 0\}$$

and ρ_{A_i}, $i \in \mathfrak{N}$, be the corresponding coherent risk measures defined in (6.252). By (6.254) with every ρ_{A_i}, $i \in \mathfrak{N}$, is associated set $\mathfrak{A}(A_i)$ of probability density functions, supported on the set A_i, such that

$$\rho_{A_i}(Z) = \sup_{\zeta\in\mathfrak{A}(A_i)} \int_\Omega \zeta(\omega) Z(\omega) dP(\omega). \tag{6.255}$$

Now let $\nu = (\nu_i)_{i\in\mathbb{N}}$ be a probability distribution (measure) on $(\Omega, \mathcal{F}_1)$, assigning probability ν_i to the event A_i, $i \in \mathbb{N}$. Assume that ν is such that $\nu(A_i) = 0$ iff $P(A_i) = 0$ (i.e., μ is absolutely continuous with respect to P and P is absolutely continuous with respect to ν on $(\Omega, \mathcal{F}_1)$); otherwise the probability measure ν is arbitrary. Define the following family of probability measures on $(\Omega, \mathcal{F}_2)$:

$$\mathfrak{C} := \left\{ \mu = \sum_{i\in\mathfrak{N}} \nu_i \mu_i : d\mu_i = \zeta_i dP,\ \zeta_i \in \mathfrak{A}(A_i),\ i \in \mathfrak{N} \right\}. \tag{6.256}$$

Note that since $\sum_{i\in\mathfrak{N}} \nu_i = 1$, every $\mu \in \mathfrak{C}$ is a probability measure. For $\mu \in \mathfrak{C}$, with respective densities $\zeta_i \in \mathfrak{A}(A_i)$ and $d\mu_i = \zeta_i dP$, and $Z \in \mathcal{Z}_2$ we have that

$$\mathbb{E}_\mu[Z|\mathcal{F}_1] = \sum_{i\in\mathfrak{N}} \mathbb{E}_{\mu_i}[Z|\mathcal{F}_1]. \tag{6.257}$$

Moreover, since ζ_i is supported on A_i,

$$\mathbb{E}_{\mu_i}[Z|\mathcal{F}_1](\omega) = \begin{cases} \int_{A_i} Z\zeta_i dP & \text{if } \omega \in A_i, \\ 0 & \text{otherwise.} \end{cases} \tag{6.258}$$

By the max-representations (6.255) it follows that for $Z \in \mathcal{Z}_2$ and $\omega \in A_i$,

$$\sup_{\mu\in\mathfrak{C}} \mathbb{E}_\mu[Z|\mathcal{F}_1](\omega) = \sup_{\zeta_i\in\mathfrak{A}(A_i)} \int_{A_i} Z\zeta_i dP = \rho_{A_i}(Z). \tag{6.259}$$

Also since $[\rho(Z)](\cdot)$ is $\mathcal{F}_1$-measurable, and hence is constant on every set A_i, we have that $[\rho(Z)](\omega) = \rho_{A_i}(Z)$ for every $\omega \in A_i$, $i \in \mathfrak{N}$. We obtain the following result.

Proposition 6.41. *Let $\mathcal{Z}_i := \mathcal{L}_p(\Omega, \mathcal{F}_i, P)$, $i = 1, 2$, with $\mathcal{F}_1 \subset \mathcal{F}_2$, and let $\rho : \mathcal{Z}_2 \to \mathcal{Z}_1$ be a conditional risk mapping. Suppose that $\mathcal{F}_1$ has a countable number of elementary events. Then*

$$\rho(Z) = \sup_{\mu\in\mathfrak{C}} \mathbb{E}_\mu[Z|\mathcal{F}_1], \quad \forall Z \in \mathcal{Z}_2, \tag{6.260}$$

where $\mathfrak{C}$ is a family of probability measures on $(\Omega, \mathcal{F}_2)$*, specified in* (6.256)*, corresponding to a probability distribution ν on* $(\Omega, \mathcal{F}_1)$.

6.7.3 Risk Averse Multistage Stochastic Programming

There are several ways in which risk averse stochastic programming can be formulated in a multistage setting. We now discuss a nested formulation similar to the derivations of section 6.7.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{F}_1 \subset \cdots \subset \mathcal{F}_T$ be a sequence of nested sigma algebras with $\mathcal{F}_1 = \{\emptyset, \Omega\}$ being trivial sigma algebra and $\mathcal{F}_T = \mathcal{F}$. (Such sequence of sigma algebras is called a *filtration*.) For $p \in [1, +\infty)$ let $\mathcal{Z}_t := \mathcal{L}_p(\Omega, \mathcal{F}_t, P)$, $t = 1, \dots, T$, be the corresponding sequence of spaces of $\mathcal{F}_t$-measurable and p-integrable functions, and let $\rho_{t+1|\mathcal{F}_t} : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$, $t = 1, \dots, T-1$, be a selected family of conditional risk mappings. It is straightforward to verify that the composition

$$\rho_{t|\mathcal{F}_{t-1}} \circ \cdots \circ \rho_{T|\mathcal{F}_{T-1}} : \mathcal{Z}_T \xrightarrow{\rho_{T|\mathcal{F}_{T-1}}} \mathcal{Z}_{T-1} \xrightarrow{\rho_{T-1|\mathcal{F}_{T-2}}} \cdots \xrightarrow{\rho_{t|\mathcal{F}_{t-1}}} \mathcal{Z}_{t-1}, \tag{6.261}$$

$t = 2, \dots, T$, of such conditional risk mappings is also a conditional risk mapping. In particular, the space $\mathcal{Z}_1$ can be identified with $\mathbb{R}$ and hence the composition $\rho_{2|\mathcal{F}_1} \circ \cdots \circ \rho_{T|\mathcal{F}_{T-1}} : \mathcal{Z}_T \to \mathbb{R}$ is a real valued coherent risk measure.

Similarly to (6.230), we consider the following nested risk averse formulation of multistage programs:

$$\begin{aligned} \operatorname*{Min}_{x_1 \in \mathcal{X}_1} f_1(x_1) + \rho_{2|\mathcal{F}_1}\Bigg[& \inf_{x_2 \in \mathcal{X}_2(x_1,\omega)} f_2(x_2, \omega) + \cdots \\ & + \rho_{T-1|\mathcal{F}_{T-2}}\Big[\inf_{x_{T-1} \in \mathcal{X}_T(x_{T-2},\omega)} f_{T-1}(x_{T-1}, \omega) \\ & + \rho_{T|\mathcal{F}_{T-1}}\Big[\inf_{x_T \in \mathcal{X}_T(x_{T-1},\omega)} f_T(x_T, \omega)\Big]\Big]\Bigg]. \end{aligned} \tag{6.262}$$

Here $f_t : \mathbb{R}^{n_{t-1}} \times \Omega \to \mathbb{R}$ and $\mathcal{X}_t : \mathbb{R}^{n_{t-1}} \times \Omega \rightrightarrows \mathbb{R}^{n_t}$, $t = 2, \dots, T$, are such that $f_t(x_t, \cdot) \in \mathcal{Z}_t$ and $\mathcal{X}_t(x_{t-1}, \cdot)$ are $\mathcal{F}_t$-measurable for all x_t and x_{t-1}.

As was discussed in section 6.7.1, the above nested formulation (6.262) has two equivalent interpretations. Namely, it can be formulated as

$$\begin{aligned} \operatorname*{Min}_{x_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T} \ & f_1(x_1) + \rho_{2|\mathcal{F}_1}\Big[f_2(\boldsymbol{x}_2(\omega), \omega) + \cdots \\ & + \rho_{T-1|\mathcal{F}_{T-2}}[f_{T-1}(\boldsymbol{x}_{T-1}(\omega), \omega) \\ & + \rho_{T|\mathcal{F}_{T-1}}[f_T(\boldsymbol{x}_T(\omega), \omega)]]\Big] \\ \text{s.t. } & x_1 \in \mathcal{X}_1,\ \boldsymbol{x}_t(\omega) \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\omega), \omega),\ t = 2, \dots, T, \end{aligned} \tag{6.263}$$

where the optimization is performed over $\mathcal{F}_t$-measurable $\boldsymbol{x}_t : \Omega \to \mathbb{R}$, $t = 1, \dots, T$, satisfying the corresponding constraints, and such that $f_t(\boldsymbol{x}_t(\cdot), \cdot) \in \mathcal{Z}_t$. Recall that the nonanticipativity is enforced here by the $\mathcal{F}_t$-measurability of $\boldsymbol{x}_t(\cdot)$. By using the *composite risk measure* $\bar{\rho} := \rho_{2|\mathcal{F}_1} \circ \cdots \circ \rho_{T|\mathcal{F}_{T-1}}$, we also can write (6.263) in the form

$$\begin{aligned} \operatorname*{Min}_{x_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T} \quad & \bar{\rho}\big[f_1(x_1) + f_2(\boldsymbol{x}_2(\omega), \omega) + \cdots + f_T(\boldsymbol{x}_T(\omega), \omega) \big] \\ \text{s.t.} \quad & x_1 \in \mathcal{X}_1,\ \boldsymbol{x}_t(\omega) \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\omega), \omega),\ t = 2, \dots, T. \end{aligned} \tag{6.264}$$

Recall that for $Z_t \in \mathcal{Z}_t$, $t = 1, \dots, T$,

$$\bar{\rho}(Z_1 + \cdots + Z_T) = Z_1 + \rho_{2|\mathcal{F}_1}\Big[Z_2 + \cdots + \rho_{T-1|\mathcal{F}_{T-2}}\big[Z_{T-1} + \rho_{T|\mathcal{F}_{T-1}}[Z_T]\big]\Big], \quad (6.265)$$

and that conditions (R$'$1)–(R$'$4) imply that $\bar{\rho} : \mathcal{Z}_T \to \mathbb{R}$ is a coherent risk measure.

Alternatively we can write the corresponding dynamic programming equations (compare with (6.236)–(6.241)):

$$Q_T(x_{T-1}, \omega) = \inf_{x_T \in \mathcal{X}_T(x_{T-1},\omega)} f_T(x_T, \omega), \quad (6.266)$$

$$Q_t(x_{t-1}, \omega) = \inf_{x_t \in \mathcal{X}_t(x_{t-1},\omega)} \Big\{ f_t(x_t, \omega) + \mathcal{Q}_{t+1}(x_t, \omega) \Big\}, \quad t = T-1, \dots, 2, \quad (6.267)$$

where

$$\mathcal{Q}_t(x_{t-1}, \omega) = \rho_{t|\mathcal{F}_{t-1}}\left[Q_t(x_{t-1}, \omega)\right], \quad t = T, \dots, 2. \quad (6.268)$$

Finally, at the first stage we solve the problem

$$\operatorname*{Min}_{x_1 \in \mathcal{X}_1} f_1(x_1) + \rho_{2|\mathcal{F}_1}[Q_2(x_1, \omega)]. \quad (6.269)$$

We need to ensure here that the cost-to-go functions are p-integrable, i.e., $Q_t(x_{t-1}, \cdot) \in \mathcal{Z}_t$ for $t = 1, \dots, T-1$ and all feasible x_{t-1}.

In applications we often deal with a data process represented by a sequence of random vectors $\xi_1, \dots, \xi_T$, say, defined on a probability space $(\Omega, \mathcal{F}, P)$. We can associate with this data process filtration $\mathcal{F}_t := \sigma(\xi_1, \dots, \xi_t)$, $t = 1, \dots, T$, where $\sigma(\xi_1, \dots, \xi_t)$ denotes the smallest sigma algebra with respect to which $\xi_{[t]} = (\xi_1, \dots, \xi_t)$ is measurable. However, it is more convenient to deal with conditional risk mappings defined directly in terms of the data process rather that the respective sequence of sigma algebras. For example, consider

$$\rho_{t|\xi_{[t-1]}}(Z) := (1-\beta_t)\mathbb{E}\left[Z|\xi_{[t-1]}\right] + \beta_t \mathsf{AV@R}_{\alpha_t}(Z|\xi_{[t-1]}), \quad t = 2, \dots, T, \quad (6.270)$$

where

$$\mathsf{AV@R}_{\alpha_t}(Z|\xi_{[t-1]}) := \inf_{Y \in \mathcal{Z}_{t-1}} \left\{ Y + \alpha_t^{-1}\mathbb{E}\big[[Z-Y]_+ \big| \xi_{[t-1]}\big] \right\}. \quad (6.271)$$

Here $\beta_t \in [0, 1]$ and $\alpha_t \in (0, 1)$ are chosen constants, $\mathcal{Z}_t := \mathcal{L}_1(\Omega, \mathcal{F}_t, P)$, where $\mathcal{F}_t$ is the smallest filtration associated with the process ξ_t, and the minimum on the right-hand side of (6.271) is taken pointwise in $\omega \in \Omega$. Compared with (6.249), the conditional AV@R is defined in (6.271) in terms of the conditional expectation with respect to the history $\xi_{[t-1]}$ of the data process rather than the corresponding sigma algebra $\mathcal{F}_{t-1}$. We can also consider conditional mean-upper-semideviation risk mappings of the form

$$\rho_{t|\xi_{[t-1]}}(Z) := \mathbb{E}[Z|\xi_{[t-1]}] + c_t \left(\mathbb{E}\Big[\big[Z - \mathbb{E}[Z|\xi_{[t-1]}]\big]_+^p \big| \xi_{[t-1]}\Big] \right)^{1/p}, \quad (6.272)$$

defined in terms of the data process. Note that with $\rho_{t|\xi_{[t-1]}}$, defined in (6.270) or (6.272), is associated coherent risk measure ρ_t which is obtained by replacing the conditional expectations with respective (unconditional) expectations. Note also that if random variable $Z \in \mathcal{Z}_t$

is independent of $\xi_{[t-1]}$, then the conditional expectations on the right-hand sides of (6.270)–(6.272) coincide with the respective unconditional expectations, and hence $\rho_{t|\xi_{[t-1]}}(Z)$ does not depend on $\xi_{[t-1]}$ and coincides with $\rho_t(Z)$.

Let us also assume that the objective functions $f_t(x_t, \xi_t)$ and feasible sets $\mathcal{X}_t(x_{t-1}, \xi_t)$ are given in terms of the data process. Then formulation (6.263) takes the form

$$\begin{aligned}\operatorname*{Min}_{x_1,\boldsymbol{x}_2,\dots,\boldsymbol{x}_T}\ & f_1(x_1) + \rho_{2|\xi_{[1]}}\Big[f_2(\boldsymbol{x}_2(\xi_{[2]}), \xi_2) + \cdots \\ & + \rho_{T-1|\xi_{[T-2]}}\Big[f_{T-1}\left(\boldsymbol{x}_{T-1}(\xi_{[T-1]}), \xi_{T-1}\right) \\ & + \rho_{T|\xi_{[T-1]}}\big[f_T\left(\boldsymbol{x}_T(\xi_{[T]}), \xi_T\right)\big]\Big]\Big] \\ \text{s.t. } & x_1 \in \mathcal{X}_1,\ \boldsymbol{x}_t(\xi_{[t]}) \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\xi_{[t-1]}), \xi_t),\ t = 2, \dots, T,\end{aligned} \tag{6.273}$$

where the optimization is performed over feasible policies.

The corresponding dynamic programming equations (6.267)–(6.268) take the form

$$Q_t(x_{t-1}, \xi_{[t]}) = \inf_{x_t \in \mathcal{X}_t(x_{t-1}, \xi_t)} \Big\{ f_t(x_t, \xi_t) + \mathcal{Q}_{t+1}(x_t, \xi_{[t]}) \Big\}, \tag{6.274}$$

where

$$\mathcal{Q}_{t+1}(x_t, \xi_{[t]}) = \rho_{t+1|\xi_{[t]}}\big[Q_{t+1}(x_t, \xi_{[t+1]})\big]. \tag{6.275}$$

Note that if the process ξ_t is stagewise independent, then the conditional expectations coincide with the respective unconditional expectations, and hence (similar to the risk neutral case) functions $\mathcal{Q}_{t+1}(x_t, \xi_{[t]}) = \mathcal{Q}_{t+1}(x_t)$ do not depend on $\xi_{[t]}$, and the cost-to-go functions $Q_t(x_{t-1}, \xi_t)$ depend only on ξ_t rather than $\xi_{[t]}$.

Of course, if we set $\rho_{t|\xi_{[t-1]}}(\cdot) := \mathbb{E}\left[\,\cdot\,|\xi_{[t-1]}\right]$, then the above equations (6.274) coincide with the corresponding risk neutral dynamic programming equations. Also, in that case the composite measure $\bar{\rho}$ becomes the corresponding expectation operator and hence formulation (6.264) coincides with the respective risk neutral formulation (3.3). Unfortunately, in the general case it is quite difficult to write the composite measure $\bar{\rho}$ in an explicit form.

Multiperiod Coherent Risk Measures

It is possible to approach risk averse multistage stochastic programming in the following framework. As before, let $\mathcal{F}_t$ be a filtration and $\mathcal{Z}_t := \mathcal{L}_p(\Omega, \mathcal{F}_t, P)$, $t = 1, \dots, T$. Consider the space $\mathcal{Z} := \mathcal{Z}_1 \times \cdots \times \mathcal{Z}_T$. Recall that since $\mathcal{F}_1 = \{\emptyset, \Omega\}$, the space $\mathcal{Z}_1$ can be identified with $\mathbb{R}$. With space $\mathcal{Z}$ we can associate its dual space $\mathcal{Z}^* := \mathcal{Z}_1^* \times \cdots \times \mathcal{Z}_T^*$, where $\mathcal{Z}_t^* = \mathcal{L}_q(\Omega, \mathcal{F}_t, P)$ is the dual of $\mathcal{Z}_t$. For $Z = (Z_1, \dots, Z_T) \in \mathcal{Z}$ and $\zeta = (\zeta_1, \dots, \zeta_T) \in \mathcal{Z}^*$ their scalar product is defined in the natural way:

$$\langle \zeta, Z \rangle := \sum_{t=1}^{T} \int_\Omega \zeta_t(\omega) Z_t(\omega) dP(\omega). \tag{6.276}$$

Note that $\mathcal{Z}$ can be equipped with a norm, consistent with $\|\cdot\|_p$ norms of its components, which makes it a Banach space. For example, we can use $\|Z\| := \sum_{t=1}^{T} \|Z_t\|_p$. This norm induces the dual norm $\|\zeta\|^* = \max\{\|\zeta_1\|_q, \dots, \|\zeta_T\|_q\}$ on the space $\mathcal{Z}^*$.

Consider a function $\varrho : \mathcal{Z} \to \mathbb{R}$. For such a function it makes sense to talk about conditions (R1), (R2), and (R4) defined in section 6.3, with $Z \succeq Z'$ understood componentwise. We say that $\varrho(\cdot)$ is a *multiperiod risk measure* if it satisfies the respective conditions (R1), (R2), and (R4). Similarly to the analysis of section 6.3, we have the following results. By Theorem 7.79 it follows from convexity (condition (R1)) and monotonicity (condition (R2)), and since $\varrho(\cdot)$ is real valued, that $\varrho(\cdot)$ is continuous. By the Fenchel–Moreau theorem, we have that convexity, continuity, and positive homogeneity (condition (R4)) imply the dual representation

$$\varrho(Z) = \sup_{\zeta \in \mathfrak{A}} \langle \zeta, Z \rangle, \quad \forall Z \in \mathcal{Z}, \tag{6.277}$$

where $\mathfrak{A}$ is a convex, bounded, and weakly* closed subset of $\mathcal{Z}^*$ (and hence, by the Banach–Alaoglu theorem, $\mathfrak{A}$ is weakly* compact). Moreover, it is possible to show, exactly in the same way as in the proof of Theorem 6.4, that condition (R2) holds iff $\zeta \succeq 0$ for every $\zeta \in \mathfrak{A}$. Conversely, if ϱ is given in the form (6.277) with $\mathfrak{A}$ being a convex weakly* compact subset of $\mathcal{Z}^*$ such that $\zeta \succeq 0$ for every $\zeta \in \mathfrak{A}$, then ϱ is a (real valued) multiperiod risk measure. An analogue of the condition (R3) (translation equivariance) is more involved; we will discuss this later.

For any multiperiod risk measure ϱ, we can formulate the risk averse multistage program

$$\begin{aligned} \operatorname*{Min}_{x_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T} \; & \varrho\big(f_1(x_1),\, f_2(\boldsymbol{x}_2(\omega), \omega), \ldots, f_T(\boldsymbol{x}_T(\omega), \omega)\big) \\ \text{s.t. } & x_1 \in \mathcal{X}_1,\; \boldsymbol{x}_t(\omega) \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\omega), \omega),\; t = 2, \ldots, T, \end{aligned} \tag{6.278}$$

where optimization is performed over $\mathcal{F}_t$-measurable $\boldsymbol{x}_t : \Omega \to \mathbb{R}$, $t = 1, \ldots, T$, satisfying the corresponding constraints, and such that $f_t(\boldsymbol{x}_t(\cdot), \cdot) \in \mathcal{Z}_t$. The nonanticipativity is enforced here by the $\mathcal{F}_t$-measurability of $\boldsymbol{x}_t(\omega)$.

Let us make the following observation. If we are currently at a certain stage of the system, then we know the past and hence it is reasonable to require that our decisions be based on that information alone and should not involve unknown data. This is the nonanticipativity constraint, which was discussed in the previous sections. However, if we believe in the considered model, we also have an idea what can and what cannot happen in the future. Think, for example, about a scenario tree representing evolution of the data process. If we are currently at a certain node of that tree, representing the current state of the system, we already know that only scenarios passing through this node can happen in the future. Therefore, apart from the nonanticipativity constraint, it is also reasonable to think about the following concept, which we refer to as the *time consistency* principle:

- At every state of the system, optimality of our decisions should not depend on scenarios which we already know cannot happen in the future.

In order to formalize this concept of time consistency we need to say, of course, what we optimize (say, minimize) at every state of the process, i.e., to formulate a respective optimality criterion associated with every state of the system. The risk neutral formulation (3.3) of multistage stochastic programming, discussed in Chapter 3, automatically satisfies the time consistency requirement (see below). The risk averse case is more involved and needs discussion. We say that multiperiod risk measure ϱ is *time consistent* if the corresponding multistage problem (6.278) satisfies the above principle of time consistency.

Consider the class of functionals $\varrho : \mathcal{Z} \to \mathbb{R}$ of the form (6.232), i.e., functionals representable as

$$\varrho(Z_1, \dots, Z_T) = Z_1 + \rho_{2|\mathcal{F}_1}\Big[Z_2 + \cdots + \rho_{T-1|\mathcal{F}_{T-2}}\big[Z_{T-1} + \rho_{T|\mathcal{F}_{T-1}}[Z_T]\big]\Big], \qquad (6.279)$$

where $\rho_{t+1|\mathcal{F}_t} : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$, $t = 1, \dots, T-1$, is a sequence of conditional risk mappings. It is not difficult to see that conditions (R$'$1), (R$'$2), and (R$'$4) (defined in section 6.7.2), applied to every conditional risk mapping $\rho_{t+1|\mathcal{F}_t}$, imply respective conditions (R1), (R2), and (R4) for the functional ϱ of the form (6.279). That is, (6.279) defines a particular class of multiperiod risk measures.

Of course, for ϱ of the form (6.279), optimization problem (6.278) coincides with the nested formulation (6.263). Recall that if the set Ω is finite, then we can formulate multistage risk averse optimization in the framework of scenario trees. As it was discussed in section 6.7.1, nested formulation (6.263) is implied by the approach where with every node of the scenario tree is associated a coherent risk measure applied to the next stage of the scenario tree. In particular, this allows us to write the corresponding dynamic programming equations and implies that an associated optimal policy has the decomposition property. That is, if the process reached a certain node at stage t, then the remaining decisions of the optimal policy are also optimal with respect to this node considered as the starting point of the process. It follows that the multiperiod risk measure of the form (6.279) is time consistent and the corresponding approach to risk averse optimization satisfies the time consistency principle.

It is interesting and important to give an intrinsic characterization of the nested approach to multiperiod risk measures. Unfortunately, this seems to be too difficult and we will give only a partial answer to this question. Let observe first that for any $Z = (Z_1, \dots, Z_T) \in \mathcal{Z}$,

$$\mathbb{E}[Z_1 + \cdots + Z_T] = Z_1 + \mathbb{E}_{|\mathcal{F}_1}\Big[Z_2 + \cdots + \mathbb{E}_{|\mathcal{F}_{T-1}}\big[Z_{T-1} + \mathbb{E}_{|\mathcal{F}_T}[Z_T]\big]\Big], \qquad (6.280)$$

where $\mathbb{E}_{|\mathcal{F}_t}[\,\cdot\,] = \mathbb{E}[\,\cdot\,|\mathcal{F}_t]$ are the corresponding conditional expectation operators. That is, the expectation risk measure $\varrho(Z_1, \dots, Z_T) := \mathbb{E}[Z_1 + \cdots + Z_T]$ is time consistent and the risk neutral formulation (3.3) of multistage stochastic programming satisfies the time consistency principle.

Consider the following condition:

(R3-d) For any $Z = (Z_1, \dots, Z_T) \in \mathcal{Z}$, $Y_t \in \mathcal{Z}_t$, $t = 1, \dots, T-1$, and $a \in \mathbb{R}$ it holds that

$$\varrho(Z_1, \dots, Z_t, Z_{t+1} + Y_t, \dots, Z_T) = \varrho(Z_1, \dots, Z_t + Y_t, Z_{t+1}, \dots, Z_T), \qquad (6.281)$$

$$\varrho(Z_1 + a, \dots, Z_T) = a + \varrho(Z_1, \dots, Z_T). \qquad (6.282)$$

Proposition 6.42. *Let $\varrho : \mathcal{Z} \to \mathbb{R}$ be a multiperiod risk measure. Then the following conditions* (i)–(iii) *are equivalent:*

(i) *There exists a coherent risk measure $\bar\rho : \mathcal{Z}_T \to \mathbb{R}$ such that*

$$\varrho(Z_1, \dots, Z_T) = \bar\rho(Z_1 + \cdots + Z_T) \quad \forall (Z_1, \dots, Z_T) \in \mathcal{Z}. \qquad (6.283)$$

(ii) *Condition* (R3-d) *is fulfilled.*

(iii) *There exists a nonempty, convex, bounded, and weakly* closed subset* $\mathfrak{A}_T$ *of probability density functions* $\mathfrak{P}_T \subset \mathcal{Z}_T^*$ *such that the dual representation* (6.277) *holds with the corresponding set* $\mathfrak{A}$ *of the form*

$$\mathfrak{A} = \big\{(\zeta_1, \dots, \zeta_T) : \zeta_T \in \mathfrak{A}_T,\ \zeta_t = \mathbb{E}[\zeta_T|\mathcal{F}_t],\ t = 1, \dots, T-1\big\}. \tag{6.284}$$

Proof. If condition (i) is satisfied, then for any $Z = (Z_1, \dots, Z_T) \in \mathcal{Z}$ and $Y_t \in \mathcal{Z}_t$,

$$\begin{aligned}\varrho(Z_1, \dots, Z_t, Z_{t+1} + Y_t, \dots, Z_T) &= \bar{\rho}(Z_1 + \cdots + Z_T + Y_t)\\ &= \varrho(Z_1, \dots, Z_t + Y_t, Z_{t+1}, \dots, Z_T).\end{aligned}$$

Property (6.282) also follows by condition (R3) of $\bar{\rho}$. That is, condition (i) implies condition (R3-d).

Conversely, suppose that condition (R3-d) holds. Then for $Z = (Z_1, Z_2, \dots, Z_T)$ we have that $\varrho(Z_1, Z_2, \dots, Z_T) = \varrho(0, Z_1 + Z_2, \dots, Z_T)$. Continuing in this way, we obtain that

$$\varrho(Z_1, \dots, Z_T) = \varrho(0, \dots, 0, Z_1 + \cdots + Z_T).$$

Define

$$\bar{\rho}(W_T) := \varrho(0, \dots, 0, W_T), \quad W_T \in \mathcal{Z}_T.$$

Conditions (R1), (R2), and (R4) for ϱ imply the respective conditions for $\bar{\rho}$. Moreover, for $a \in \mathbb{R}$ we have

$$\begin{aligned}\bar{\rho}(W_T + a) &= \varrho(0, \dots, 0, W_T + a) = \varrho(0, \dots, a, W_T) = \cdots = \varrho(a, \dots, 0, W_T)\\ &= a + \varrho(0, \dots, 0, W_T) = \bar{\rho}(W_T) + a.\end{aligned}$$

That is, $\bar{\rho}$ is a coherent risk measure, and hence (ii) implies (i).

Now suppose that condition (i) holds. By the dual representation (see Theorem 6.4 and Proposition 6.5), there exists a convex, bounded, and weakly* closed set $\mathfrak{A}_T \subset \mathfrak{P}_T$ such that

$$\bar{\rho}(W_T) = \sup_{\zeta_T \in \mathfrak{A}_T} \langle \zeta_T, W_T \rangle, \quad W_T \in \mathcal{Z}_T. \tag{6.285}$$

Moreover, for $W_T = Z_1 + \cdots + Z_T$ we have $\langle \zeta_T, W_T \rangle = \sum_{t=1}^T \mathbb{E}\left[\zeta_T Z_t\right]$, and since Z_t is $\mathcal{F}_t$-measurable,

$$\mathbb{E}\left[\zeta_T Z_t\right] = \mathbb{E}\big[\mathbb{E}[\zeta_T Z_t|\mathcal{F}_t]\big] = \mathbb{E}\big[Z_t \mathbb{E}[\zeta_T|\mathcal{F}_t]\big]. \tag{6.286}$$

That is, (i) implies (iii). Conversely, suppose that (iii) holds. Then (6.285) defines a coherent risk measure $\bar{\rho}$. The dual representation (6.277) together with (6.284) imply (6.283). This shows that conditions (i) and (iii) are equivalent. ☐

As we know, condition (i) of the above proposition is necessary for the multiperiod risk measure ϱ to be representable in the nested form (6.279). (See section 6.7.3 and equation (6.265) in particular.) This condition, however, is not sufficient. It seems to be quite difficult to give a complete characterization of coherent risk measures $\bar{\rho}$ representable in the form

$$\bar{\rho}(Z_1 + \cdots + Z_T) = Z_1 + \rho_{2|\mathcal{F}_1}\Big[Z_2 + \cdots + \rho_{T-1|\mathcal{F}_{T-2}}\big[Z_{T-1} + \rho_{T|\mathcal{F}_{T-1}}[Z_T]\big]\Big] \tag{6.287}$$

for all $Z = (Z_1, \dots, Z_T) \in \mathcal{Z}$, and some sequence $\rho_{t+1|\mathcal{F}_t} : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$, $t = 1, \dots, T-1$, of conditional risk mappings.

Remark 26. Of course, condition $\zeta_t = \mathbb{E}[\zeta_T | \mathcal{F}_t]$, $t = 1, \dots, T-1$, of (6.284) can be written as

$$\zeta_t = \mathbb{E}[\zeta_{t+1} | \mathcal{F}_t], \quad t = 1, \dots, T-1. \tag{6.288}$$

That is, if representation (6.283) holds for some coherent risk measure $\bar{\rho}(\cdot)$, then any element $(\zeta_1, \dots, \zeta_T)$ of the dual set $\mathfrak{A}$, in the representation (6.277) of $\varrho(\cdot)$, forms a *martingale* sequence.

Example 6.43. Let $\rho_{\tau|\mathcal{F}_{\tau-1}} : \mathcal{Z}_\tau \to \mathcal{Z}_{\tau-1}$ be a conditional risk mapping for some $2 \le \tau \le T$, and let $\rho_1(Z_1) := Z_1$, $Z_1 \in \mathbb{R}$, and $\rho_{t|\mathcal{F}_{t-1}} := \mathbb{E}_{|\mathcal{F}_{t-1}}$, $t = 2, \dots, T$, $t \ne \tau$. That is, we take here all conditional risk mappings to be the respective conditional expectations except (an arbitrary) conditional risk mapping $\rho_{\tau|\mathcal{F}_{\tau-1}}$ at the period $t = \tau$. It follows that

$$\begin{aligned} \varrho(Z_1, \dots, Z_T) &= \mathbb{E}\left[Z_1 + \cdots + Z_{\tau-1} + \rho_{\tau|\mathcal{F}_{\tau-1}}\left[\mathbb{E}_{|\mathcal{F}_\tau}[Z_\tau + \cdots + Z_T]\right]\right] \\ &= \mathbb{E}\left[\rho_{\tau|\mathcal{F}_{\tau-1}}\left[\mathbb{E}_{|\mathcal{F}_\tau}[Z_1 + \cdots + Z_T]\right]\right]. \end{aligned} \tag{6.289}$$

That is,

$$\bar{\rho}(W_T) = \mathbb{E}\left[\rho_{\tau|\mathcal{F}_{\tau-1}}[\mathbb{E}_{|\mathcal{F}_\tau}[W_T]]\right], \quad W_T \in \mathcal{Z}_T, \tag{6.290}$$

is the corresponding (composite) coherent risk measure.

Coherent risk measures of the form (6.290) have the following property:

$$\bar{\rho}(W_T + Y_{\tau-1}) = \bar{\rho}(W_T) + \mathbb{E}[Y_{\tau-1}], \quad \forall W_T \in \mathcal{Z}_T,\ \forall Y_{\tau-1} \in \mathcal{Z}_{\tau-1}. \tag{6.291}$$

By (6.284) the above condition (6.291) means that the corresponding set $\mathfrak{A}$, defined in (6.284), has the additional property that $\zeta_t = \mathbb{E}[\zeta_T] = 1$, $t = 1, \dots, \tau-1$, i.e., these components of $\zeta \in \mathfrak{A}$ are constants (equal to one).

In particular, for $\tau = T$ the composite risk measure (6.290) becomes

$$\bar{\rho}(W_T) = \mathbb{E}\left[\rho_{T|\mathcal{F}_{T-1}}[W_T]\right], \quad W_T \in \mathcal{Z}_T. \tag{6.292}$$

Further, let $\rho_{T|\mathcal{F}_{T-1}} : \mathcal{Z}_T \to \mathcal{Z}_{T-1}$ be the conditional mean absolute deviation, i.e.,

$$\rho_{T|\mathcal{F}_{T-1}}[Z_T] := \mathbb{E}_{|\mathcal{F}_{T-1}}\Big[Z_T + c\left|Z_T - \mathbb{E}_{|\mathcal{F}_{T-1}}[Z_T]\right|\Big], \tag{6.293}$$

$c \in [0, 1/2]$. The corresponding composite coherent risk measure here is

$$\bar{\rho}(W_T) = \mathbb{E}[W_T] + c\,\mathbb{E}\left|W_T - \mathbb{E}_{|\mathcal{F}_{T-1}}[W_T]\right|, \quad W_T \in \mathcal{Z}_T. \tag{6.294}$$

For $T > 2$ the risk measure (6.294) is different from the mean absolute deviation measure

$$\tilde{\rho}(W_T) := \mathbb{E}[W_T] + c\,\mathbb{E}\big|W_T - \mathbb{E}[W_T]\big|, \quad W_T \in \mathcal{Z}_T, \tag{6.295}$$

and that the multiperiod risk measure

$$\varrho(Z_1, \dots, Z_T) := \tilde{\rho}(Z_1 + \cdots + Z_T) = \mathbb{E}[Z_1 + \cdots + Z_T] + c\,\mathbb{E}\big|Z_1 + \cdots + Z_T - \mathbb{E}[Z_1 + \cdots + Z_T]\big|$$

corresponding to (6.295) is not time consistent. ■

Risk Averse Multistage Portfolio Selection

We discuss now the example of portfolio selection. A nested formulation of multistage portfolio selection can be written as

$$\begin{array}{ll} \text{Min} & \left\{\bar{\rho}(-W_T) := \rho_1\Big[\cdots\rho_{T-1|W_{T-2}}\big[\rho_{T|W_{T-1}}[-W_T]\big]\Big]\right\} \\ \text{s.t.} & W_{t+1} = \sum_{i=1}^n \xi_{i,t+1}x_{it},\ \sum_{i=1}^n x_{it} = W_t,\ x_t \geq 0,\ t = 0,\cdots,T-1. \end{array} \tag{6.296}$$

We use here conditional risk mappings formulated in terms of the respective conditional expectations, like the conditional AV@R (see (6.270)) and conditional mean semideviations (see (6.272)), and the notation $\rho_{t|W_{t-1}}$ stands for a conditional risk mapping defined in terms of the respective conditional expectations given W_{t-1}. By $\rho_t(\cdot)$ we denote the corresponding (unconditional) risk measures. For example, to the conditional $\mathsf{AV@R}_\alpha(\,\cdot\,|\xi_{[t-1]})$ corresponds the respective (unconditional) $\mathsf{AV@R}_\alpha(\,\cdot\,)$. If we set $\rho_{t|W_{t-1}} := \mathbb{E}_{|W_{t-1}}$, $t = 1,\dots,T$, then since

$$\mathbb{E}\Big[\cdots\mathbb{E}\big[\mathbb{E}[-W_T|W_{T-1}]\,\big|W_{T-2}\big]\Big] = \mathbb{E}[-W_T],$$

we obtain the risk neutral formulation. Note also that in order to formulate this as a minimization, rather than a maximization, problem we changed the sign of ξ_{it}.

Suppose that the random process ξ_t is *stagewise independent*. Let us write dynamic programming equations. At the last stage we have to solve problem

$$\begin{array}{ll} \underset{x_{T-1}\geq 0,\,W_T}{\text{Min}} & \rho_{T|W_{T-1}}[-W_T] \\ \text{s.t.} & W_T = \sum_{i=1}^n \xi_{iT}x_{i,T-1},\ \sum_{i=1}^n x_{i,T-1} = W_{T-1}. \end{array} \tag{6.297}$$

Since W_{T-1} is a function of $\xi_{[T-1]}$, by the stagewise independence we have that ξ_T, and hence W_T, are independent of W_{T-1}. It follows by positive homogeneity of ρ_T that the optimal value of (6.297) is $Q_{T-1}(W_{T-1}) = W_{T-1}\nu_{T-1}$, where ν_{T-1} is the optimal value of

$$\begin{array}{ll} \underset{x_{T-1}\geq 0,\,W_T}{\text{Min}} & \rho_T[-W_T] \\ \text{s.t.} & W_T = \sum_{i=1}^n \xi_{iT}x_{i,T-1},\ \sum_{i=1}^n x_{i,T-1} = 1, \end{array} \tag{6.298}$$

and an optimal solution of (6.297) is $\bar{x}_{T-1}(W_{T-1}) = W_{T-1}x^*_{T-1}$, where x^*_{T-1} is an optimal solution of (6.298). Continuing in this way, we obtain that the optimal policy $\bar{x}_t(W_t)$ here is *myopic*. That is, $\bar{x}_t(W_t) = W_t x^*_t$, where x^*_t is an optimal solution of

$$\begin{array}{ll} \underset{x_t\geq 0,\,W_{t+1}}{\text{Min}} & \rho_{t+1}[-W_{t+1}] \\ \text{s.t.} & W_{t+1} = \sum_{i=1}^n \xi_{i,t+1}x_{it},\ \sum_{i=1}^n x_{it} = 1 \end{array} \tag{6.299}$$

(compare with section 1.4.3). Note that the composite risk measure $\bar{\rho}$ can be quite complicated here.

An alternative, multiperiod risk averse approach can be formulated as

$$\begin{aligned}&\text{Min } \rho[-W_T]\\&\text{s.t. } W_{t+1}=\sum_{i=1}^{n}\xi_{i,t+1}x_{it},\ \sum_{i=1}^{n}x_{it}=W_t,\ x_t\ge 0,\ t=0,\dots,T-1,\end{aligned} \tag{6.300}$$

for an explicitly defined risk measure ρ. Let, for example,

$$\rho(\cdot):=(1-\beta)\mathbb{E}[\,\cdot\,]+\beta\mathsf{AV@R}_\alpha(\,\cdot\,),\quad \beta\in[0,1],\ \alpha\in(0,1). \tag{6.301}$$

Then problem (6.300) becomes

$$\begin{aligned}&\text{Min } (1-\beta)\mathbb{E}[-W_T]+\beta\big(-r+\alpha^{-1}\mathbb{E}[r-W_T]_+\big)\\&\text{s.t. } W_{t+1}=\sum_{i=1}^{n}\xi_{i,t+1}x_{it},\ \sum_{i=1}^{n}x_{it}=W_t,\ x_t\ge 0,\ t=0,\dots,T-1,\end{aligned} \tag{6.302}$$

where $r\in\mathbb{R}$ is the (additional) first-stage decision variable. After r is decided, at the first stage, the problem comes to minimizing $\mathbb{E}[U(W_T)]$ at the last stage, where $U(W):=(1-\beta)W+\beta\alpha^{-1}[W-r]_+$ can be viewed as a disutility function.

The respective dynamic programming equations become as follows. The last-stage value function $Q_{T-1}(W_{T-1},r)$ is given by the optimal value of the problem

$$\begin{aligned}&\underset{x_{T-1}\ge 0,\,W_T}{\text{Min}}\ \mathbb{E}\big[-(1-\beta)W_T+\beta\alpha^{-1}[r-W_T]_+\big]\\&\text{s.t. } W_T=\sum_{i=1}^{n}\xi_{iT}x_{i,T-1},\ \sum_{i=1}^{n}x_{i,T-1}=W_{T-1}.\end{aligned} \tag{6.303}$$

Proceeding in this way, at stages $t=T-2,\dots,1$ we consider the problems

$$\begin{aligned}&\underset{x_t\ge 0,\,W_{t+1}}{\text{Min}}\ \mathbb{E}\{Q_{t+1}(W_{t+1},r)\}\\&\text{s.t. } W_{t+1}=\sum_{i=1}^{n}\xi_{i,t+1}x_{it},\ \sum_{i=1}^{n}x_{it}=W_t,\end{aligned} \tag{6.304}$$

whose optimal value is denoted $Q_t(W_t,r)$. Finally, at stage $t=0$ we solve the problem

$$\begin{aligned}&\underset{x_0\ge 0,r,W_1}{\text{Min}}\ -\beta r+\mathbb{E}[Q_1(W_1,r)]\\&\text{s.t. } W_1=\sum_{i=1}^{n}\xi_{i1}x_{i0},\ \sum_{i=1}^{n}x_{i0}=W_0.\end{aligned} \tag{6.305}$$

In the above multiperiod risk averse approach, the optimal policy is not myopic and the property of time consistency is not satisfied.

Risk Averse Multistage Inventory Model

Consider the multistage inventory problem (1.17). The nested risk averse formulation of that problem can be written as

$$
\begin{aligned}
\operatorname*{Min}_{x_t \ge y_t}\ & c_1(x_1 - y_1) + \rho_1\Big[\psi_1(x_1, D_1) + c_2(x_2 - y_2) + \rho_{2|D_{[1]}}\big[\psi_2(x_2, D_2) + \cdots \\
& \quad + c_{T-1}(x_{T-1} - y_{T-1}) + \rho_{T-1|D_{[T-2]}}\big[\psi_{T-1}(x_{T-1}, D_{T-1}) \\
& \quad + c_T(x_T - y_T) + \rho_{T|D_{[T-1]}}[\psi_T(x_T, D_T)]\big]\big]\Big] \\
\text{s.t. }\ & y_{t+1} = x_t - D_t,\ t = 1, \dots, T-1,
\end{aligned}
\tag{6.306}
$$

where y_1 is a given initial inventory level, $\psi_t(x_t, d_t) := b_t[d_t - x_t]_+ + h_t[x_t - d_t]_+$, and $\rho_{t|D_{[t-1]}}(\cdot)$, $t = 2, \dots, T$, are chosen conditional risk mappings. Recall that the notation $\rho_{t|D_{[t-1]}}(\cdot)$ stands for a conditional risk mapping obtained by using conditional expectations, conditional on $D_{[t-1]}$, and note that $\rho_1(\cdot)$ is real valued and is a coherent risk measure.

As discussed earlier, there are two equivalent interpretations of problem (6.306). We can write it as an optimization problem with respect to feasible policies $\boldsymbol{x}_t(d_{[t-1]})$ (compare with (6.273)):

$$
\begin{aligned}
\operatorname*{Min}_{x_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T}\ & c_1(x_1 - y_1) + \rho_1\Big[\psi_1(x_1, D_1) + c_2(\boldsymbol{x}_2(D_1) - x_1 + D_1) \\
& + \rho_{2|D_1}\big[\psi_2(\boldsymbol{x}_2(D_1), D_2) + \cdots \\
& + c_{T-1}(\boldsymbol{x}_{T-1}(D_{[T-2]}) - \boldsymbol{x}_{T-2}(D_{[T-3]}) + D_{T-2}) \\
& + \rho_{T-1|D_{[T-2]}}\big[\psi_{T-1}(\boldsymbol{x}_{T-1}(D_{[T-2]}), D_{T-1}) \\
& + c_T(\boldsymbol{x}_T(D_{[T-1]}) - \boldsymbol{x}_{T-1}(D_{[T-2]}) + D_{T-1}) \\
& + \rho_{T|D_{[T-1]}}[\psi_T(\boldsymbol{x}_T(D_{[T-1]}), D_T)]\big]\big]\Big] \\
\text{s.t. }\ & x_1 \ge y_1,\ \boldsymbol{x}_2(D_1) \ge x_1 - D_1, \\
& \boldsymbol{x}_t(D_{[t-1]}) \ge \boldsymbol{x}_{t-1}(D_{[t-2]}) - D_{t-1},\ t = 3, \dots, T.
\end{aligned}
\tag{6.307}
$$

Alternatively, we can write dynamic programming equations. At the last stage $t = T$, for observed inventory level y_T, we need to solve the problem

$$
\operatorname*{Min}_{x_T \ge y_T}\ c_T(x_T - y_T) + \rho_{T|D_{[T-1]}}\big[\psi_T(x_T, D_T)\big]. \tag{6.308}
$$

The optimal value of problem (6.308) is denoted $Q_T(y_T, D_{[T-1]})$. Continuing in this way, we write for $t = T-1, \dots, 2$ the following dynamic programming equations:

$$
Q_t(y_t, D_{[t-1]}) = \min_{x_t \ge y_t} c_t(x_t - y_t) + \rho_{t|D_{[t-1]}}\big[\psi(x_t, D_t) + Q_{t+1}\big(x_t - D_t, D_{[t]}\big)\big]. \tag{6.309}
$$

Finally, at the first stage we need to solve the problem

$$
\operatorname*{Min}_{x_1 \ge y_1}\ c_1(x_1 - y_1) + \rho_1\big[\psi(x_1, D_1) + Q_2(x_1 - D_1, D_1)\big]. \tag{6.310}
$$

Suppose now that the process D_t is stagewise independent. Then, by exactly the same argument as in section 1.2.3, the cost-to-go (value) function $Q_t(y_t, d_{[t-1]}) = Q_t(y_t)$, $t = 2, \dots, T$, is independent of $d_{[t-1]}$, and by convexity arguments the optimal policy $\bar{x}_t = \bar{\boldsymbol{x}}_t(d_{[t-1]})$ is a basestock policy. That is, $\bar{x}_t = \max\{y_t, x_t^*\}$, where x_t^* is an optimal solution of

$$\operatorname*{Min}_{x_t} c_t x_t + \rho_t\big[\psi(x_t, D_t) + Q_{t+1}(x_t - D_t)\big]. \tag{6.311}$$

Recall that ρ_t denotes the coherent risk measure corresponding to the conditional risk mapping $\rho_{t|D_{[t-1]}}$.

Exercises

6.1. Let $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$ be a random variable with cdf $H(z) := P\{Z \le z\}$. Note that $\lim_{z\downarrow t} H(z) = H(t)$ and denote $H^-(t) := \lim_{z\uparrow t} H(z)$. Consider functions $\phi_1(t) := \mathbb{E}[t - Z]_+$, $\phi_2(t) := \mathbb{E}[Z - t]_+$ and $\phi(t) := \beta_1\phi_1(t) + \beta_2\phi_2(t)$, where β_1, β_2 are positive constants. Show that ϕ_1, ϕ_2, and ϕ are real valued convex functions with subdifferentials

$$\partial\phi_1(t) = [H^-(t), H(t)] \text{ and } \partial\phi_2(t) = [-1 + H^-(t), -1 + H(t)],$$

$$\partial\phi(t) = [(\beta_1 + \beta_2)H^-(t) - \beta_2, (\beta_1 + \beta_2)H(t) - \beta_2].$$

Conclude that the set of minimizers of $\phi(t)$ over $t \in \mathbb{R}$ is the (closed) interval of $[\beta_2/(\beta_1 + \beta_2)]$-quantiles of $H(\cdot)$.

6.2. (i) Let $Y \sim \mathcal{N}(\mu, \sigma^2)$. Show that

$$\mathsf{V@R}_\alpha(Y) = \mu + z_\alpha\sigma, \tag{6.312}$$

where $z_\alpha := \Phi^{-1}(1 - \alpha)$, and

$$\mathsf{AV@R}_\alpha(Y) = \mu + \frac{\sigma}{\alpha\sqrt{2\pi}} e^{-z_\alpha^2/2}. \tag{6.313}$$

(ii) Let $Y^1, \dots, Y^N$ be an iid sample of $Y \sim \mathcal{N}(\mu, \sigma^2)$. Compute the asymptotic variance and asymptotic bias of the sample estimator $\hat{\theta}_N$, of $\theta^* = \mathsf{AV@R}_\alpha(Y)$, defined on page 300.

6.3. Consider the chance constraint

$$\Pr\left\{\sum_{i=1}^n \xi_i x_i \ge b\right\} \ge 1 - \alpha, \tag{6.314}$$

where $\xi \sim \mathcal{N}(\mu, \Sigma)$ (see problem (1.43)). Note that this constraint can be written as

$$\mathsf{V@R}_\alpha\left(b - \sum_{i=1}^n \xi_i x_i\right) \le 0. \tag{6.315}$$

Consider the following constraint:

$$\mathsf{AV@R}_\gamma\left(b - \sum_{i=1}^{n} \xi_i x_i\right) \le 0. \tag{6.316}$$

Show that constraints (6.314) and (6.316) are equivalent if $z_\alpha = \frac{1}{\gamma\sqrt{2\pi}} e^{-z_\gamma^2/2}$.

6.4. Consider the function $\phi(x) := \mathsf{AV@R}_\alpha(F_x)$, where $F_x = F_x(\omega) = F(x,\omega)$ is a real valued random variable, on a probability space $(\Omega, \mathcal{F}, P)$, depending on $x \in \mathbb{R}^n$. Assume that (i) for a.e. $\omega \in \Omega$ the function $F(\cdot,\omega)$ is continuously differentiable on a neighborhood V of a point $x_0 \in \mathbb{R}^n$, (ii) the families $|F(x,\omega)|$, $x \in V$, and $\|\nabla_x F(x,\omega)\|$, $x \in V$, are dominated by a P-integrable function, and (iii) the random variable F_x has continuous distribution for all $x \in V$. Show that under these conditions, $\phi(x)$ is directionally differentiable at x_0 and

$$\phi'(x_0, d) = \alpha^{-1} \inf_{t\in[a,b]} \mathbb{E}\left\{ d^{\mathsf{T}} \nabla_x([F(x_0,\omega) - t]_+)\right\}, \tag{6.317}$$

where a and b are the respective left- and right-side $(1-\alpha)$-quantiles of the cdf of the random variable F_{x_0}. Conclude that if, moreover, $a = b = \mathsf{V@R}_\alpha(F_{x_0})$, then $\phi(\cdot)$ is differentiable at x_0 and

$$\nabla\phi(x_0) = \alpha^{-1}\mathbb{E}\left[\mathbf{1}_{\{F_{x_0} > a\}}(\omega)\nabla_x F(x_0,\omega)\right]. \tag{6.318}$$

Hint: Use Theorem 7.44 together with the Danskin theorem, Theorem 7.21.

6.5. Show that the set of saddle points of the minimax problem (6.190) is given by $\{\mu\} \times [\gamma^*, \gamma^{**}]$, where γ^* and γ^{**} are defined in (6.192).

6.6. Consider the absolute semideviation risk measure

$$\rho_c(Z) := \mathbb{E}\{Z + c[Z - \mathbb{E}(Z)]_+\}, \quad Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P),$$

where $c \in [0,1]$, and the following risk averse optimization problem:

$$\operatorname*{Min}_{x\in X} \underbrace{\mathbb{E}\left\{G(x,\xi) + c[G(x,\xi) - \mathbb{E}(G(x,\xi))]_+\right\}}_{\rho_c[G(x,\xi)]}. \tag{6.319}$$

Viewing the optimal value of problem (6.319) as the Von Mises statistical functional of the probability measure P, compute its influence function.

Hint: Use derivations of section 6.5.3 together with the Danskin theorem.

6.7. Consider the risk averse optimization problem (6.162) related to the inventory model. Let the corresponding risk measure be of the form $\rho_\lambda(Z) = \mathbb{E}[Z] + \lambda\mathbb{D}(Z)$, where $\mathbb{D}(Z)$ is a measure of variability of $Z = Z(\omega)$ and λ is a nonnegative trade-off coefficient between expectation and variability. Higher values of λ reflect a higher degree of risk aversion. Suppose that ρ_λ is a coherent risk measure for all $\lambda \in [0,1]$ and let S_λ be the set of optimal solutions of the corresponding risk averse problem. Suppose that the sets S_0 and S_1 are nonempty.

Show that if $S_0 \cap S_1 = \emptyset$, then S_λ is monotonically nonincreasing or monotonically nondecreasing in $\lambda \in [0,1]$ depending on whether $S_0 > S_1$ or $S_0 < S_1$. If $S_0 \cap S_1 \ne \emptyset$, then $S_\lambda = S_0 \cap S_1$ for any $\lambda \in (0,1)$.

6.8. Consider the news vendor problem with cost function

$$F(x,d) = cx + b[d-x]_+ + h[x-d]_+, \quad \text{where } b > c \geq 0,\ h > 0,$$

and the minimax problem

$$\operatorname*{Min}_{x\geq 0} \sup_{H\in\mathfrak{M}} \mathbb{E}_H[F(x,D)], \tag{6.320}$$

where $\mathfrak{M}$ is the set of cumulative distribution functions (probability measures) supported on (final) interval $[l,u] \subset \mathbb{R}_+$ and having a given mean $\bar{d} \in [l,u]$. Show that for any $x \in [l,u]$ the maximum of $\mathbb{E}_H[F(x,D)]$ over $H \in \mathfrak{M}$ is attained at the probability measure $\bar{H} = p\Delta(l) + (1-p)\Delta(u)$, where $p = (u-\bar{d})/(u-l)$, i.e., the cdf $\bar{H}(\cdot)$ is the step function

$$\bar{H}(z) = \begin{cases} 0 & \text{if} \quad z < l, \\ p & \text{if} \quad l \leq z < u, \\ 1 & \text{if} \quad u \leq z. \end{cases}$$

Conclude that $\bar{H}$ is the cdf specified in Proposition 6.38 and that $\bar{x} = \bar{H}^{-1}(\kappa)$, where $\kappa = (b-c)/(b+h)$, is the optimal solution of problem (6.320). That is, $\bar{x} = l$ if $\kappa < p$ and $\bar{x} = u$ if $\kappa > p$, where $\kappa = \frac{b-c}{b+h}$.

6.9. Consider the following version of the news vendor problem. A news vendor has to decide about quantity x of a product to purchase at the cost of c per unit. He can sell this product at the price s per unit and unsold products can be returned to the vendor at the price of r per unit. It is assumed that $0 \leq r < c < s$. If the demand D turns out to be greater than or equal to the order quantity x, then he makes profit $sx - cx = (s-c)x$, while if D is less than x, his profit is $sD + r(x-D) - cx$. Thus the profit is a function of x and D and is given by

$$F(x,D) = \begin{cases} (s-c)x & \text{if} \quad x \leq D, \\ (r-c)x + (s-r)D & \text{if} \quad x > D. \end{cases} \tag{6.321}$$

(a) Assuming that demand $D \geq 0$ is a random variable with cdf $H(\cdot)$, show that the expectation function $f(x) := \mathbb{E}_H[F(x,D)]$ can be represented in the form

$$f(x) = (s-c)x - (s-r)\int_0^x H(z)dz. \tag{6.322}$$

Conclude that the set of optimal solutions of the problem

$$\operatorname*{Max}_{x\geq 0} \left\{ f(x) := \mathbb{E}_H[F(x,D)] \right\} \tag{6.323}$$

is an interval given by the set of κ-quantiles of the cdf $H(\cdot)$ with $\kappa := (s-c)/(s-r)$.

(b) Consider the following risk averse version of the news vendor problem:

$$\operatorname*{Min}_{x\geq 0} \left\{ \phi(x) := \rho[-F(x,D)] \right\}. \tag{6.324}$$

Here ρ is a real valued coherent risk measure representable in the form (6.165) and H^* is the corresponding reference cdf.

(i) Show that the function $\phi(x) = \rho[-F(x, D)]$ can be represented in the form

$$\phi(x) = (c - s)x + (s - r) \int_0^x \bar{H}(z)dz \tag{6.325}$$

for some cdf $\bar{H}$.

(ii) Show that if $\rho(\cdot) := \mathsf{AV@R}_\alpha(\cdot)$, then $\bar{H}(z) = \max\left\{\alpha^{-1}H^*(z), 1\right\}$. Conclude that in that case, optimal solutions of the risk averse problem (6.324) are smaller than the risk neutral problem (6.323).

6.10. Let $\mathcal{Z}_i := \mathcal{L}_p(\Omega, \mathcal{F}_i, P)$, $i = 1, 2$, with $\mathcal{F}_1 \subset \mathcal{F}_2$, and let $\rho : \mathcal{Z}_2 \to \mathcal{Z}_1$.

(a) Show that if ρ is a conditional risk mapping, $Y \in \mathcal{Z}_1$ and $Y \succeq 0$, then $\rho(YZ) = Y\rho(Z)$ for any $Z \in \mathcal{Z}_2$.

(b) Suppose that the mapping ρ satisfies conditions (R$'$1)–(R$'$3), but not necessarily the positive homogeneity condition (R$'$4). Show that it can be represented in the form

$$[\rho(Z)](\omega) = \sup_{\mu \in \mathfrak{C}} \left\{\mathbb{E}_\mu[Z|\mathcal{F}_1](\omega) - [\rho^*(\mu)](\omega)\right\}, \tag{6.326}$$

where $\mathfrak{C}$ is a set of probability measures on $(\Omega, \mathcal{F}_2)$ and

$$[\rho^*(\mu)](\omega) = \sup_{Z \in \mathcal{Z}_2} \left\{\mathbb{E}_\mu[Z|\mathcal{F}_1](\omega) - [\rho(Z)](\omega)\right\}. \tag{6.327}$$

You may assume that $\mathcal{F}_1$ has a countable number of elementary events.

6.11. Consider the following risk averse approach to multistage portfolio selection. Let $\xi_1, \ldots, \xi_T$ be the respective data process (of random returns) and consider the following chance constrained nested formulation:

$$\begin{aligned}
&\text{Max } \mathbb{E}[W_T] \\
&\text{s.t. } W_{t+1} = \sum_{i=1}^n \xi_{i,t+1} x_{it}, \;\; \sum_{i=1}^n x_{it} = W_t, \;\; x_t \geq 0, \\
&\qquad \mathsf{Pr}\left\{W_{t+1} \geq \kappa W_t \,\middle|\, \xi_{[t]}\right\} \geq 1 - \alpha, \;\; t = 0, \ldots, T-1,
\end{aligned} \tag{6.328}$$

where $\kappa \in (0, 1)$ and $\alpha \in (0, 1)$ are given constants. Dynamic programming equations for this problem can be written as follows. At the last stage $t = T - 1$, the cost-to-go function $Q_{T-1}(W_{T-1}, \xi_{[T-1]})$ is given by the optimal value of the problem

$$\begin{aligned}
&\operatorname*{Max}_{x_{T-1} \geq 0, W_T} \; \mathbb{E}\left[W_T \,\middle|\, \xi_{[T-1]}\right] \\
&\text{s.t. } W_T = \sum_{i=1}^n \xi_{iT} x_{i,T-1}, \;\; \sum_{i=1}^n x_{i,T-1} = W_{T-1}, \\
&\qquad \mathsf{Pr}\left\{W_T \geq \kappa W_{T-1} \,\middle|\, \xi_{[T-1]}\right\},
\end{aligned} \tag{6.329}$$

and at stage $t = T-2, \dots, 1$, the cost-to-go function $Q_t(W_t, \xi_{[t]})$ is given by the optimal value of the problem

$$\begin{array}{ll} \operatorname*{Max}_{x_t \geq 0,\, W_{t+1}} & \mathbb{E}\big[Q_{t+1}(W_{t+1}, \xi_{[t+1]}) \,\big|\, \xi_{[t]}\big] \\ \text{s.t.} & W_{t+1} = \displaystyle\sum_{i=1}^{n} \xi_{i,t+1} x_{i,t}, \ \sum_{i=1}^{n} x_{i,t} = W_t, \\ & \Pr\big\{W_{t+1} \geq \kappa W_t \,\big|\, \xi_{[t]}\big\}. \end{array} \tag{6.330}$$

Assuming that the process ξ_t is stagewise independent, show that the optimal policy is myopic and is given by $\bar{x}_t(W_t) = W_t x_t^*$, where x_t^* is an optimal solution of the problem

$$\begin{array}{ll} \operatorname*{Max}_{x_t \geq 0} & \displaystyle\sum_{i=1}^{n} \mathbb{E}\left[\xi_{i,t+1}\right] x_{i,t} \\ \text{s.t.} & \displaystyle\sum_{i=1}^{n} x_{i,t} = 1, \ \Pr\left\{\sum_{i=1}^{n} \xi_{i,t+1} x_{i,t} \geq \kappa\right\} \geq 1-\alpha. \end{array} \tag{6.331}$$