INFORMS TutORials in Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Coherent Approaches to Risk in Optimization Under Uncertainty

R. Tyrrell Rockafellar

To cite this entry: R. Tyrrell Rockafellar. Coherent Approaches to Risk in Optimization Under Uncertainty. *In* INFORMS TutORials in Operations Research. Published online: 14 Oct 2014; 38-61.
https://doi.org/10.1287/educ.1073.0032

TUTORIALS IN
OPERATIONS RESEARCH
INFORMS 2007

 | ISBN 13 978-1-877640-22-3
DOI 10.1287/educ.1073.0032

Coherent Approaches to Risk in Optimization Under Uncertainty

R. Tyrrell Rockafellar
Department of Industrial and Systems Engineering, University of Florida, Gainesville, Florida 32611, rtr@ise.ufl.edu

Abstract Decisions often need to be made before all the facts are in. A facility must be built to withstand storms, floods, or earthquakes of magnitudes that can only be guessed from historical records. A portfolio must be purchased in the face of only statistical knowledge, at best, about how markets will perform. In optimization, this implies that constraints may need to be envisioned in terms of safety margins instead of exact requirements. But what does that really mean in model formulation? What guidelines make sense, and what are the consequences for optimization structure and computation?

The idea of a coherent measure of risk in terms of surrogates for potential loss, which has been developed in recent years for applications in financial engineering, holds promise for a far wider range of applications in which the traditional approaches to uncertainty have been subject to criticism. The general ideas and main facts are presented here with the goal of facilitating their transfer to practical work in those areas.

Keywords optimization under uncertainty; safeguarding against risk; safety margins; measures of risk; measures of potential loss; measures of deviation; coherency; value-at-risk; conditional value-at-risk; probabilistic constraints; quantiles; risk envelopes; dual representations; stochastic programming

1. Introduction

In classical optimization based on deterministic modeling, a typical problem in n variables has the form

$$\text{minimize } c_0(x) \quad \text{over all } x \in S \text{ satisfying } c_i(x) \le 0 \text{ for } i = 1, \ldots, m, \tag{1.1}$$

where S is a subset of $\mathbb{R}^n$ composed of vectors $x = (x_1, \ldots, x_n)$, and each c_i is a function from S to $\mathbb{R}$. In the environment of uncertainty that dominates a vast range of applications, however, a serious difficulty arises in such a formulation. We can think of it as caused by parameter elements about which the optimizer (who wishes to solve the problem) has only incomplete information at the time x must be chosen. Decisions are then fraught with risk over their outcomes, and the way to respond may be puzzling.

The difficulty can be captured by supposing that instead of just $c_i(x)$ we have $c_i(x, \omega)$, where ω belongs to a set Ω representing future states of knowledge. For instance, Ω might be a subset of some parameter space $\mathbb{R}^d$, or merely a finite index set. The choice of an $x \in S$ no longer produces specific numbers $c_i(x)$, as taken for granted in problem (1.1), but merely results in a collection of *functions* on Ω

$$\underline{c}_i(x)\colon \omega \to c_i(x, \omega) \quad \text{for } i = 0, 1, \ldots, m. \tag{1.2}$$

How then should the constraints in the problem be construed? How should the objective be reinterpreted? In what ways should risk be taken into account? Safeguards may be needed

to protect against undesired outcomes, and safety margins may have to be introduced, but on what basis?

Various approaches, with their pros and cons, have commonly been followed and will be reviewed shortly as background for explaining more recent ideas. However, an important principle should be understood first: No conceptual distinction should be made between the treatment of the objective function c_0 and the constraint functions $c_1, \ldots, c_m$ in these circumstances.

Behind the formulation of problem (1.1) there may have been a number of functions, all of interest in terms of keeping their values low. A somewhat arbitrary choice may have been made as to which one should be minimized subject to constraints on the others. Apart from this arbitrariness, well known device, in which an additional coordinate x_{n+1} is appended to $x = (x_1, \ldots, x_n)$, can anyway always be used to convert (1.1) into an equivalent problem in which all the complications are put into the constraints and none are left in the objective, namely

$$\begin{aligned} &\text{minimize } x_{n+1} \quad \text{over all } (x, x_{n+1}) \in S \times \mathbb{R} \\ &\text{satisfying } \begin{cases} c_0(x) - x_{n+1} \le 0 & \text{and} \\ c_i(x) \le 0 \quad \text{for } i = 1, \ldots, m. \end{cases} \end{aligned} \tag{1.3}$$

The challenges of uncertainty would be faced in the reformulated model with the elements $\omega \in \Omega$ affecting only constraints.

This principle will help in seeing the modeling implications of different approaches to handling risk. Let us also note, before proceeding further, that in the notation $c_i(x, \omega)$ some functions might only depend on a partial aspect of ω, or perhaps not on ω at all, although for our purposes, constraints not touched by uncertainty could be suppressed into the specification of the set S. Equations have been omitted as constraints because, if uncertain, they rarely make sense in this basic setting, and if certain, they could likewise be put into S.

Hereafter, we will think of Ω as having the mathematical structure of a probability space with a probability measure P for comparing the likelihood of future states ω. This is more a technical device rather than a philosophical statement. Perhaps there is true knowledge of probabilities, or a subjective belief in probabilities that should appropriately influence actions by the decision maker. But the theory to be described here will bring into consideration other probability measures as alternatives, and ways of suggesting at least what the probabilities of subdivisions of Ω might be if a more complete knowledge is lacking. The designation P might therefore be just a way to begin.

By working with a probability measure P on Ω we can interpret the functions $\underline{c}_i(x)\colon \Omega \to \mathbb{R}$ as *random variables*. Any function $X\colon \Omega \to \mathbb{R}$ induces a probability distribution on $\mathbb{R}$ with cumulative distribution function F_X defined by taking $F_X(z)$ to be the probability assigned by P to the set of $\omega \in \Omega$ such that $X(\omega) \le z$. (In the theory of probability spaces introducing a field of measurable sets in Ω, and so forth, should be a concern. For this tutorial, however, such details are not considered.)

Integrals with respect to P will be written as expectations E. We will limit attention to random variables X for which $E[X^2]$ is finite; such random variables make up a linear space that will be denoted here just by $\mathcal{L}^2$. Having $X \in \mathcal{L}^2$ ensures that both the mean and standard deviation of X, namely

$$\mu(X) = EX \qquad \text{and} \qquad \sigma(X) = (E[(X - \mu(X))^2])^{1/2}$$

are well defined and finite. Here we will assume that all the specific random variables entering our picture through (1.2) for choices of $x \in S$ belong to $\mathcal{L}^2$.

To recapitulate, in this framework uncertainty causes the numbers $c_i(x)$ in the deterministic model (1.1) to be replaced by the random variables $\underline{c}_i(x)$ in (1.2). This casts doubt on interpretation of the constraints and objective. Some remodeling is therefore required before a problem of optimization is again achieved.

2. Some Traditional Approaches

Much can be said about how to address uncertainty in optimization, and how it should affect the modeling done in a specific application. But, the most fundamental idea is to begin by condensing random variables that depend on x back to numbers that depend on x. We will discuss several of the most familiar ways of doing this and compare their features. In the next section, a broader perspective will be taken and a theory furnishing guidelines will be developed.

In coming examples, we adopt the same approach in each case to the objective and every constraint, although approaches could be mixed in practice. This will underscore the principle of not thinking that constraints require different modes of treatment than objectives. It will also help to clarify shortcomings in these approaches.

2.1. Approach 1: Guessing the Future

A common approach in practice, serving essentially as a way of avoiding the issues, is to identify a single element $\bar{\omega} \in \Omega$ as furnishing a best estimate of the unknown information, and then to

$$\text{minimize } c_0(x, \bar{\omega}) \quad \text{over all } x \in S \text{ satisfying } c_i(x, \bar{\omega}) \leq 0 \text{ for } i = 1, \ldots, m. \tag{2.1}$$

Although this might be justifiable when the uncertainty is minor and well concentrated around $\bar{\omega}$, it is otherwise subject to serious criticism. A solution $\bar{x}$ to (2.1) could lead, when the future state turns out to be some ω other than $\bar{\omega}$, to a constraint value $c_i(\bar{x}, \omega) > 0$, or a cost $c_0(\bar{x}, \omega)$ disagreeably higher than $c_0(\bar{x}, \bar{\omega})$. No provision has been made for the *risk* inherent in these eventualities. A decision $\bar{x}$ coming out of (2.1) fails to hedge against the uncertainty and thus "puts all the eggs in one basket." It does not incorporate any appraisal of how harmful an ultimate constraint violation or cost overrun might be to the application being modeled.

The weakness in this response to uncertainty can also be appreciated from another angle. If Ω has been modeled as a continuum in a space $\mathbb{R}^d$ of parameter vectors, the behavior of solutions to (2.1) as an optimization problem depending on $\bar{\omega}$ as a parameter element could be poor. Even in linear programming it is well understood that tiny changes in coefficients can produce big changes, even jumps, in solutions. The dangers of not hedging could be seriously compounded by such instability.

2.2. Approach 2: Worst-Case Analysis

Another familiar approach is to rely on determining the worst that might happen. In its purest form, the counterpart to the deterministic problem (1.1) obtained in this way is to

$$\text{minimize } \sup_{\omega \in \Omega} c_0(x, \omega) \quad \text{over all } x \in S \text{ satisfying } \sup_{\omega \in \Omega} c_i(x, \omega) \leq 0 \text{ for } i = 1, \ldots, m. \tag{2.2}$$

Here we write "sup" instead of "max" not only to avoid claiming attainment by some ω, but also in deference to the technicality that we must be dealing with the essential least upper bound (neglecting sets of probability 0) when Ω is infinite.

This very conservative formulation aims at ensuring that the constraints will be satisfied, no matter what the future brings. It devotes attention only to the worst possible outcomes, even if they are associated only with future states thought to be highly unlikely. Assessments of performance in more ordinary cirumstances are not addressed.

Although the goal is to eliminate *all* risk, there is a price for that. The feasible set, consisting of the xs satisfying all the constraints, might be very small—possibly empty. For example, if the uncertainty in ω has to do with storms, floods, or earthquakes, and x is tied to the design of a structure intended to withstand these forces, there may be no available choice of x guaranteeing absolute compliance. The optimizer may have to live with a balance between the practicality of x and the chance that the resulting design could be overpowered by some extreme event.

Nonetheless a strong attraction of this formulation is that the potential trouble over specifying a probability measure P on Ω is effectively bypassed. A modern and more sophisticated version of the worst-case approach, motivated by that feature, is currently promoted as *robust optimization*. It aims likewise to avoid the introduction of a probability measure, but tries anyway to treat some parts of Ω as more important (more likely) than others. A generalization along these lines will be given as Approach 8 in §6.4.

2.3. Approach 3: Relying on Expectations

Still another idea of long standing, back in the context of Ω being a probability space, is to utilize the expectations of the random variables $\underline{c}_i(x)$ as numbers that depend on x. Taking this approach at its purest, one could

$$\text{minimize } E[\underline{c}_0(x)] \quad \text{over all } x \in S \text{ satisfying } E[\underline{c}_i(x)] \le 0 \text{ for } i = 1, \ldots, m. \tag{2.3}$$

As far as the objective is concerned, this is a normal way of proceeding, and it has a long history. Yet for the constraints it seems generally ridiculous. If a constraint corresponded to the safety of a structure, for example, or the avoidance of bankruptcy, who would be satisfied with it only being fulfilled on the average? Expectations are primarily suitable for situations where the interest lies in long-range operation, and where stochastic ups and downs can safely average out. To the contrary, many applications have a distinctly short-run focus with serious risks in the foreground.

Why then should the expectation approach be acceptable for the objective in (2.3)? That runs counter to the no-distinction-in-treatment principle explained in the introduction. More will be seen about this below.

2.4. Approach 4: Standard Deviation Units as Safety Margins

An appealing way to dramatically improve on expectation constraints is to introduce safety margins based on standard deviation so as to ensure that the expected value is not just 0 but reassuringly below 0. For a choice of positive values $\lambda_i > 0$, the constraints set up in this manner take the form

$$\mu(\underline{c}_i(x)) + \lambda_i \sigma(\underline{c}_i(x)) \le 0 \quad \text{for } i = 1, \ldots, m. \tag{2.4}$$

The significance is that the future states ω for which one gets $c_i(x, \omega) > 0$ instead of the desired $c_i(x, \omega) \le 0$ correspond only to the upper part of the distribution of the random variable $\underline{c}_i(x)$ that lies more than λ_i standard deviation units above the expected value of $\underline{c}_i(x)$. This resonates with many stipulations in statistical estimation about levels of confidence. Furthermore, it affords a compromise with the harsh conservatism of the worst-case approach.

What is the comparable formulation to (2.4) to adopt for the objective? The answer is to introduce another coefficient $\lambda_0 > 0$ and

$$\text{minimize } \mu(\underline{c}_0(x)) + \lambda_0 \sigma(\underline{c}_0(x)) \quad \text{over all } x \in S \text{ satisfying (2.4).} \tag{2.5}$$

This is an interesting way to look at the objective, though it is almost never considered. Its interpretation, along the lines of (1.3) with the objective values viewed as costs is that one is looking for the lowest level of x_{n+1} as a cost threshold such that, for some $x \in S$ satisfying (2.4), the cost outcomes $c_0(x, \omega) > x_{n+1}$ will occur only in states ω corresponding to the high end of the distribution of $\underline{c}_i(x)$ lying more than λ_0 standard deviation units above the mean cost.

Despite the seeming simplicity and attractiveness of this idea, it has a major flaw that will stand out when we get into the theoretical consideration of what guidelines should prevail for good modeling. A key property called coherency is lacking. A powerful substitute without this defect is presented in Approach 9 in §7.4.

2.5. Approach 5: Specifying Probabilities of Compliance

Another popular alternative to the worst-case approach, and which bears some resemblence to the one just outlined, is to pass to *probabilistic* constraints (also called *chance* constraints) in which the desired inequalities $c_i(x,\omega) \le 0$ are to hold at least with specified probabilities

$$\text{prob}\{\underline{c}_i(x) \le 0\} \ge \alpha_i, \quad \text{for } i = 1,\dots,m, \tag{2.6}$$

where α_i is a confidence level, say 0.99. Following the idea in (1.3) for handling the objective, the problem is to

$$\begin{aligned} &\text{minimize } x_{n+1} \text{ over all } (x, x_{n+1}) \in S \times \mathbb{R} \text{ satisfying} \\ &\qquad \text{prob}\{\underline{c}_0(x) \le x_{n+1}\} \ge \alpha_0 \text{ and the constraints (2.6).} \end{aligned} \tag{2.7}$$

For instance, with $\alpha_0 = 0.5$ one would be choosing x to get the *median* of the random variable $\underline{c}_0(x)$, rather than its mean value, as low as possible.

Drawbacks are found even in this mode of optimization modeling, however. A qualitative objection, like the one about relying on a confidence level specified by standard deviation units, is that inadequate account is taken of the degree of danger inherent in potential violations beyond that level. In the cases where constraint violations $c_i(x,\omega) > 0$ occur, which they do with probability $1 - \alpha_i$, is there merely inconvenience or a disaster? The specification of α_i, alone, does not seem to fully address that. A technical objection too, from the optimization side, is that the probability expressions in (2.6) and (2.7) can exhibit poor mathematical behavior with respect to x, often lacking convexity and even continuity.

It is less apparent that Approach 5, like its predecessors, fits the pattern of condensing a random variable into a single number. Yet it does—in terms of quantiles and *value-at-risk*, a central idea in finance. In (2.6), the α_i-quantile of the random variable $\underline{c}_i(x)$ must be ≤ 0, but technicalities can make the precise meaning problematic. This is discussed further in §5 when we explain value-at-risk and some recent approaches with *conditional value-at-risk* and its variants involving risk profiles.

2.6. Constraint Consolidation

Questions could be raised about the appropriateness of the tactic in Approaches 4 and 5 of putting a separate probability-dependent condition on each random variable. Why not, for instance, put $\underline{c}_1(x),\dots,\underline{c}_m(x)$ into a single random variable

$$\underline{c}(x) \text{ with } c(x,\omega) = \max\big\{c_1(x,\omega),\dots,c_m(x,\omega)\big\} \tag{2.8}$$

and then constrain $\text{prob}\{\underline{c}(x) \le 0\} \ge \alpha$? That would correspond to insisting that x be feasible with probability at least α. Nothing here should be regarded as counseling against that idea, which is very reasonable. However, the consequences may not be as simple as imagined.

The units in which the different costs $c_i(x,\omega)$ are presented may be quite different. Issues of scaling could arise with implications for the behavior of a condition on $\underline{c}(x)$ alone. Should each $\underline{c}_i(x)$ be multiplied first by some $\lambda_i > 0$ in (2.8) to adjust to this? If so, how should these coefficients be chosen?

Note also that individual constraints like those in (2.4) or (2.6) allow some costs to be subjected to tighter control than others, which is lost when they are consolidated into a single cost. A combination of individual constraints and a consolidated constraint may be appropriate. Not to be overlooked either is the no-distinction-in-treatment principle for the objective. But how should the objective be brought in?

Having raised these issues, we now put them in the background and continue with separate conditions on the random variables $\underline{c}_i(x)$. This theory will anyway be applicable to alternative formulations involving constraint consolidation.

2.7. Stochastic Programming and Multistage Futures

Stochastic programming is a major area of methodology dedicated to optimization problems under uncertainty (Wets [22]). Its leading virtue is extended modeling of the future, especially through *recourse decisions*. Simply put, instead of choosing $x \in S$ and then having to cope with its consequences when a future state $\omega \in \Omega$ is reached, there may be an opportunity then for a second decision x' that could counteract bad consequences, or take advantage of good consequences of x. There could then be passage to a state ω' further in the future, and perhaps yet another decision x'' after that. Although we will not go into this important subject, we note that the newer ideas explained here have yet to be incorporated in stochastic programming. When applied to our bare-bones format of choosing x and then experiencing ω, the traditional formulation in stochastic programming would be to minimize the expectation of $\underline{c}_0(x)$ over all $x \in S$ satisfying $\sup \underline{c}_i(x) \le 0$ for $i = 1, \ldots, m$. In particular, the objective and the constraints are treated quite differently. It should be clear from the comments about Approaches 2 and 3 that the improvements may need to be considered. Theoretical work in that direction has been initiated in Ruszczyński and Shapiro [21]. However, it should also be understood that stochastic programming strives, as far as possible in the modeling process, to eliminate uncertain constraints through the possibilites for recourse and various penalty expressions that might relate them. The purpose is not so much to obtain an exact solution as it is to identify ways of hedging that might otherwise be overlooked. The themes about risk which we develop here could assist further in that effort.

2.8. Dynamic Programming

Dynamic programming is another area of methodology in optimization under uncertainty that focuses on a future with many stages—perhaps an infinite number of stages. Dynamic programming operates backward in time to the present. Because it is more concerned with policies for controlling an uncertain system than coping with near-term risk, it is outside of the scope of this tutorial.

2.9. Penalty Staircases

A common and often effective approach to replacing the simple minimization of $c_0(x)$ by something involving the random variable $\underline{c}_0(x)$, when uncertainty sets in, without merely passing to $E[\underline{c}_0(x)]$, is to

$$\text{minimize } E[\psi(\underline{c}_0(x))] \quad \text{for an increasing convex function } \psi \text{ on } (-\infty, \infty). \tag{2.9}$$

For example, a series of cost thresholds $d_1, \ldots, d_q$ might could be specified, and ψ could be taken to be a piecewise linear function having breakpoints at $d_1, \ldots, d_q$, which imposes increasingly steeper penalization rates as successive thresholds are exceeded. A drawback is the difficulty in predicting how the selection of ψ will shape the distribution of $\underline{c}_0(x)$ in optimality.

An alternative would be to proceed in the manner of Approach 5 with the choice of objective but to supplement it with constraints such as

$$\text{prob}\{\underline{c}_0^k(x) - d_k \le 0\} \ge \alpha_0^k, \quad \text{for } k = 1, \ldots, q. \tag{2.10}$$

The point is that a random variable like $\underline{c}_0(x)$ can be propagated into a sequence of other random variables

$$\underline{c}_0^k(x) = \underline{c}_0(x) - d_k \quad \text{for } k = 1, \ldots, q \tag{2.11}$$

in staircase fashion to achieve sharper control over results. Of course, the probabilistic constraints in (2.10) are only a temporary proposal, given the defects already discussed. Better ways of dealing with a *staircased random variable* as in (2.11) will soon be available.

3. Quantification of Risk

We wish to paint a large-scale picture of risk, without being bound to any one viewpoint or area of application, and to supply an axiomatic foundation that assists in clearing up some of the persistent misunderstandings.

What is risk? Everyone agrees that risk is associated with having to make a decision without fully knowing its consequences, due to future uncertainty, but also knowing that some of those consequences might be bad, or at least undesirable relative to others. Still, how might the *quantity* of risk be evaluated to construct a model in which optimization can be carried out?

Two basic ideas must be considered and coordinated. To many people, the amount of risk in a random variable representing a cost of some kind is the degree of *uncertainty* in it, i.e., how much it deviates from being constant. To other people, risk must be quantified in terms of a surrogate for the overall cost, such as its mean value, median value, or worst possible value. All the examples surveyed so far in optimization under uncertainty have revolved around such surrogates, but both ways of viewing risk will have roles in this tutorial.

The meaning of "cost" can be very general: Money, pollution, underperformance, safety hazard, failure to meet an obligation, etc. In optimization the concern is often a cost that is relative to some target and keeping it below 0, so that it does not become a "loss." Of course, a negative cost or loss amounts to a "gain."

For clarity, we will speak of *measures of deviation* when assessing inconstancy, with the standard deviation of a random variable serving as terminological inspiration. We will speak of *measures of risk* when assigning a single value to a random variable as a surrogate for its overall cost. Although this conforms to current common usage, it seems to create a competition between the second kind of measure and the first. It would really be more accurate to speak of the second kind as measures of the risk *of loss*, so we will use that terminology initially, before reverting to just speaking of measures of risk, for short.

Random variables could represent many things, but to achieve our goal of handling the random variables $\underline{c}_i(x)$ in (1.2) that come out of an optimization problem such as (1.1) when uncertainty clouds the formulation, it is important to adopt an orientation. When speaking of a measure of risk of loss being applied to a random variable X, we will always have in mind that X represents a cost, as above: Positive outcomes $X(\omega)$ of X are disliked, and large positive outcomes disliked even more, while negative outcomes are welcomed. This corresponds with traditions in optimization in which quantities are typically minimized or constrained to be ≤ 0.

The core of the difficulty in optimization under uncertainty is the fact that a random variable is not, itself, a single quantity. The key to coping with this will be to condense the random variable into a single quantity by *quantifying the risk of loss*, rather than the degree of uncertainty, in it. We are thinking of positive outcomes of random variables $\underline{c}_i(x)$ associated with constraints as losses (e.g., cost overruns). This presupposes that variables have been set up so that constraints are in ≤ 0 form. For $\underline{c}_0(x)$, associated with minimization, loss is unnecessary cost incurred by not making the best choices.

The random variables X in our framework are identified with functions from Ω to $\mathbb{R}$ that belong to the linear space $\mathcal{L}^2$ which we introduced relative to a probability measure P on Ω. They can be added, multiplied by scalars, and so forth. In quantifying loss, we must assign to each $X \in \mathcal{L}^2$ a value $\mathcal{R}(X)$. We will take this value to belong to $(-\infty, \infty]$. In addition to having it be a real number, we may allow it in some circumstances to be ∞. Quantifying risk of loss will therefore revolve around specifying a functional $\mathcal{R}$ from $\mathcal{L}^2$ to $(-\infty, \infty]$. (In mathematics, a function on a space of functions, like $\mathcal{L}^2$, is typically called a functional.)

3.1. A General Approach to Uncertainty in Optimization

In the context of problem (1.1) disrupted by uncertainty causing the function values $c_i(x)$ to be replaced by the random variables $\underline{c}_i(x)$ in (1.2), select for each $i = 0, 1, \ldots, m$ a functional

$\mathcal{R}_i$: $\mathcal{L}^2 \to (-\infty, \infty]$ aimed at quantifying the risk of loss. Then

$$\begin{gathered}\text{replace the random variables } \underline{c}_i(x) \text{ by the functions } \bar{c}_i(x) = \mathcal{R}_i(\underline{c}_i(x)) \text{ and} \\ \text{minimize } \bar{c}_0(x) \text{ over all } x \in S \text{ satisfying } \bar{c}_i(x) \le 0 \text{ for } i = 1, \ldots, m. \end{gathered} \tag{3.1}$$

As an important variant, $\underline{c}_0(x)$ could be *staircased* in the manner of (2.11), and the same could be done for $\underline{c}_1(x), \ldots, \underline{c}_m(x)$, if desired. This would introduce a multiplicity of random variables $\underline{c}_i^k(x)$ that could individually be composed with functionals $\mathcal{R}_i^k$ to supplement (3.1) with constraints providing additional control over the results of optimization.

Because the end product of any staircasing would still look like problem (3.1) except in notation, it will suffice, in the theory, to deal with (3.1).

The fundamental question now is this: What axiomatic properties should a functional have to be a good quantifier of the risk of loss? The pioneering work of Artzner et al. [3, 4], Delbaen [6], has provided a solid answer in terms of properties they identified as providing *coherency*. Their work concentrated on applications in finance, especially banking regulations, but the contribution goes far beyond that. We will present the concept in a form that follows more recent developments in expanding the ideas of those authors, or in some respects simplifying or filling in details. The differences will be discussed after the definition.

It will be convenient to use C to stand for either a number in $\mathbb{R}$ or the corresponding constant random variable $X \equiv C$ in $\mathcal{L}^2$. We write $X \le X'$ meaning that $X(\omega) \le X'(\omega)$ with probability 1, and so forth. The *magnitude* of an $X \in \mathcal{L}^2$ is given by the norm

$$\|X\|_2 = (E[X^2])^{1/2} = (\mu^2(X) + \sigma^2(X))^{1/2}. \tag{3.2}$$

A sequence of random variables X^k, $k = 1, 2, \ldots$, *converges* to a random variable X with respect to this norm if $\|X^k - X\|_2 \to 0$, or equivalently, if both $\mu(X^k - X) \to 0$ and $\sigma(X^k - X) \to 0$ as $k \to \infty$.

3.2. Coherent Measures of Risk

A functional $\mathcal{R}$: $\mathcal{L}^2 \to (-\infty, \infty]$ will be called a *coherent measure of risk in the extended sense* if

(R1) $\mathcal{R}(C) = C$ for all constants C,
(R2) $\mathcal{R}((1-\lambda)X + \lambda X') \le (1-\lambda)\mathcal{R}(X) + \lambda\mathcal{R}(X')$ for $\lambda \in (0,1)$ ("convexity"),
(R3) $\mathcal{R}(X) \le \mathcal{R}(X')$ when $X \le X'$ ("monotonicity"),
(R4) $\mathcal{R}(X) \le 0$ when $||X^k - X||_2 \to 0$ with $\mathcal{R}(X^k) \le 0$ ("closedness").

It will be called a *coherent measure of risk in the basic sense* if it also satisfies

(R5) $\mathcal{R}(\lambda X) = \lambda\mathcal{R}(X)$ for $\lambda > 0$ ("positive homogeneity").

The original definition of coherency in Artzner et al. [3, 4] required (R5). Insistence on this scaling property has since been called into question on various fronts. Although it is the dropping of it that we have in mind in supplementing their basic definition by an extended one, there are also other, lesser, differences between this version and the original definition.

Property (R1), which implies in particular that $\mathcal{R}(0) = 0$, has the motivation that if a random variable always has the same outcome C, then in condensing it to a single surrogate value, that value ought to be C. In Artzner et al. [3, 4] the place of (R1) was taken by a more complicated property, which was tailored to a banking concept, but came down to having

$$\mathcal{R}(X + C) = \mathcal{R}(X) + C \quad \text{for constants } C. \tag{3.3}$$

This extension of (R1) follows *automatically* from the combination of (R1) and (R2), as was shown in Rockafellar et al. [16]. In that paper, however, as well as in Artzner et al. [3, 4] the orientation was different: Random variables were viewed not as costs but as anticosts

(affording gains or rewards), which would amount to a switch of sign, transforming (3.3) into $\mathcal{R}(X+C) = \mathcal{R}(X) - C$ and (R1) into $\mathcal{R}(C) = -C$. The formulation in this tutorial is dictated by the wish for a straightforward adaptation to the conventions of optimization theory, so that recent developments about risk can be at home in that subject and can smoothly reach a wider audience.

The risk inequality in (R3) has similarly been affected by the switch in orientation from anticosts to costs. The same will be true for a number of other conditions and formulas discussed or cited below, although this will not always be mentioned.

The combination of (R2) with (R5) leads to *subadditivity*

$$\mathcal{R}(X + X') \leq \mathcal{R}(X) + \mathcal{R}(X'). \tag{3.4}$$

On the other hand this property together with (R5) implies (R2). Subadditivity was emphasized in Artzner et al. [3, 4] as a key property that was lacking in the approaches popular with practioners in finance. The interpretation is that when X and X' are the loss variables (cost = loss) for two different portfolios, the total risk of loss should be reduced, or a least not made worse, when the portfolios are combined into one. This refers to *diversification*. The same argument can be offered as justification for the more basic property of convexity in (R2). Forming a weighted mixture of two portfolios should not increase overall loss potential. Otherwise there might be something to be gained by partioning a portfolio (or comparable entity outside of finance) into increasingly smaller fractions.

The monotonicity in (R3) also makes perfect sense. If $X(\omega) \leq X'(\omega)$ almost surely in the future states ω, the risk of loss seen in X should not exceed the risk of loss seen in X', with respect to quantification by $\mathcal{R}$. Yet some seemingly innocent strategies among practioners violate this, as will be seen shortly.

Note from applying (R3) in the case of $X' = \sup X$, when X is bounded from above, and invoking (R1), that

$$\mathcal{R}(X) \leq \sup X \quad \text{always}, \tag{3.5}$$

and, on the other hand, from taking $X' = 0$ instead, that

$$\mathcal{R}(X) \leq 0 \quad \text{when } X \leq 0. \tag{3.6}$$

The latter property is in fact *equivalent* to the monotonicity in (R3) through the convexity in (R2).

3.3. Acceptable Risks

Artzner et al. [3, 4] introduced the terminology that the risk (of loss) associated with a random variable X is *acceptable* with respect to a choice of a coherent risk measure $\mathcal{R}$ when $\mathcal{R}(X) \leq 0$. This ties in with the idea that $\mathcal{R}(X)$ is the surrogate being adopted for the potential cost or loss. By (3.6), X is acceptable in particular if it exhibits no chance of producing a positive cost. But the concept allows compromise where there could sometimes be positive costs, as long as they are not overwhelming by some carefully chosen standard. The examples discussed in the next section explain this in greater detail.

In this respect, axiom (R4) says that if a random variable X can be approximated arbitrarily closely by acceptable random variables X^k, then X too should be acceptable. Such an approximation axiom was not included in the original papers of Artzner et al. [3, 4], but that may have been due to the fact that in those papers Ω was a finite set and the finiteness of $\mathcal{R}(X)$ was taken for granted. Because a finite convex function on a finite-dimensional space is automatically continuous, the closedness in (R5) would then be automatic. It has been noted by Ruszczyński that the continuity of $\mathcal{R}$ follows also for infinite-dimensional $\mathcal{L}^2$ from the combination of (R2) and (R3) as long as $\mathcal{R}$ does not take on ∞. But ∞ values can occur in some examples which we do not wish to exclude.

3.4. Coherency in Optimization

The main consequences of coherency in adopting (3.1) as a basic model for optimization under uncertainty are summarized as follows:

Theorem 1. *Suppose in problem* (3.1)*, posed with functions* $\bar{c}_i(x) = \mathcal{R}_i(\underline{c}_i(x))$ *for* $i = 0, 1, \ldots, m$*, that each functional* $\mathcal{R}_i$ *is a coherent measures of risk in the extended sense.*

(a) Preservation of convexity. *If* $c_i(x, \omega)$ *is convex with respect to* x *for each* ω*, then the function* $\bar{c}_i(x) = \mathcal{R}_i(\underline{c}_i(x))$ *is convex. Thus, if problem* (1.1) *without uncertainty would have been a problem of convex programming, that advantage persists when uncertainty enters and is handled by passing to the formulation in* (3.1) *with coherency.*

(b) Preservation of certainty. *If* $\underline{c}_i(x)$ *is actually just a constant random variable for each* x*, i.e.,* $c_i(x, \omega) = c_i(x)$ *with no influence from* ω*, then* $\bar{c}_i(x) = c_i(x)$*. Thus, the composition technique does not distort problem features that were not subject to uncertainty.*

(c) Insensitivity to scaling. *If the risk measures* $\mathcal{R}_i$ *also satisfy* (R5)*, then problem* (3.1) *remains the same when the units in which the values* $c_i(x, \omega)$ *are denominated are rescaled.*

Property (a) of Theorem 1 holds through (R2) and (R3) because the composition of a convex function with a *nondecreasing* convex function is another convex function. (Without (R3), this could definitely fail.) Property (b) is immediate from (R1), whereas (c) corresponds to (R5).

Note that the constraints $\bar{c}_i(x) \le 0$ in problem (3.1) correspond to requiring x to *make the risk in the random variable* $\underline{c}_i(x)$ *be acceptable* according to the dictates of the selected risk measure $\mathcal{R}_i$. A related feature coming out of (3.3), is that

$$\mathcal{R}_i(\underline{c}_i(x)) \le b_i \iff \mathcal{R}_i(\underline{c}_i(x) - b_i) \le 0. \tag{3.7}$$

Thus, acceptability is stable under translation and does not depend on where the zero is located in the scale of units for a random variable.

4. Coherency or Its Lack in Traditional Approaches

It is time to return to the traditional approaches to see how coherent they may or may not be. We will also look at the standards they implicitly adopt for deeming the risk in a cost random variable to be acceptable.

4.1. Approach 1: Guessing the Future

This corresponds to assessing the risk in $\underline{c}_i(x)$ as $\mathcal{R}(\underline{c}_i(x))$ with

$$\mathcal{R}(X) = X(\bar{\omega}) \text{ for some choice of } \bar{\omega} \in \Omega \text{ having positive probability.} \tag{4.1}$$

This functional $\mathcal{R}$ does give a coherent measure of risk in the basic sense, but is open to criticism if used for such a purpose in responding to uncertainty. The risk in X is regarded as acceptable if there is no positive cost in the future state $\bar{\omega}$. No account is taken of any other future states.

4.2. Approach 2: Worst-Case Analysis

This corresponds to assessing the risk in $\underline{c}_i(x)$ as $\mathcal{R}(\underline{c}_i(x))$ with

$$\mathcal{R}(X) = \sup X. \tag{4.2}$$

Again, we have a coherent measure of risk in the basic sense, but it is severely conservative. Note that this furnishes an example where perhaps $\mathcal{R}(X) = \infty$. That will happen

whenever X does not have a finite upper bound (almost surely), which for finite Ω, could not happen. The risk in X is acceptable only when $X \leq 0$, so that positive costs have zero probability.

4.3. Approach 3: Relying on Expectations

This corresponds to assessing the risk in $\underline{c}_i(x)$ as $\mathcal{R}(\underline{c}_i(x))$ with

$$\mathcal{R}(X) = \mu(X) = EX. \tag{4.3}$$

This is a coherent measure of risk in the basic sense, but it is feeble. Acceptability of the risk in X merely refers to negative costs being enough to balance out positive costs in the long run.

4.4. Approach 4: Standard Deviation Units as Safety Margins

This corresponds to assessing the risk in $\underline{c}_i(x)$ as $\mathcal{R}_i(\underline{c}_i(x))$ with

$$\mathcal{R}_i(X) = \mu(X) + \lambda_i \sigma(X) \quad \text{for some } \lambda_i > 0. \tag{4.4}$$

However, such a functional $\mathcal{R}_i$ is *not* a coherent measure of risk. Axiom (R3) fails, although (R1), (R2), (R4) and even (R5) hold. This is one of the prime examples that the authors in Artzner et al. [3, 4] had in mind when developing their concept of coherency, because it lies at the heart of classical approaches to risk in finance.

Note that because (R3) fails for (4.4), the introduction of safety margins in this manner can *destroy convexity* when forming composites as in problem (3.1), and thus eliminate the benefits in part (a) of Theorem 1. This is unfortunate. Acceptability of the risk in X means that positive costs can only occur in the part of the distribution of X that lies more than λ_i standard deviation units above the mean. However, an excellent substitute that preserves convexity will emerge below in terms of conditional value-at-risk and other versions of safety margins based on various other measures of deviation.

4.5. Approach 5: Specifying Probabilities of Compliance

This corresponds to assessing the risk in $\underline{c}_i(x)$ as $\mathcal{R}_i(\underline{c}_i(x))$ with

$$\mathcal{R}_i(X) = q_{\alpha_i}(X) = \alpha_i\text{-quantile in the distribution of } X, \text{ for a choice of } \alpha_i \in (0,1). \tag{4.5}$$

Although the precise meaning will be explained in the next section, it must be noted that this does *not* furnish a coherent measure of risk. The difficulty here lies in the convexity axiom (R2), which is equivalent to the combination of the positive homogeneity in (R5) and the subadditivity in (3.4). Although (R5) is obeyed, the subadditivity in (3.4), standing for the desirability of diversification, is violated. This was another important motivation for the development of coherency in Artzner et al. [3, 4]. Quantiles correspond in finance to value-at-risk, which is even incorporated into international banking regulations.

Without coherency, this approach, like the one before it, can *destroy convexity* that might otherwise be available for optimization modeling. Convexity can be salvaged in (4.5) if the distributions of the random variables $\underline{c}_i(x)$ belong to the *log-concave* class for all $x \in S$, but even then there are technical hurdles because the convexity of $\mathcal{R}_i$ is missing relative to the entire space $\mathcal{L}^2$.

For (4.5) acceptability of the risk in X means, of course, that positive costs are avoided with probability α_i. Again, this is a natural idea. Although the faults in it are dismaying, *conditional* value-at-risk will address them.

5. Value-at-Risk and Conditional Value-at-Risk

In terms of the cumulative distribution function F_X of a random variable X and a probability level $\alpha \in (0,1)$, the *value-at-risk* $\mathrm{VaR}_\alpha(X)$ and the *α-quantile* $q_\alpha(X)$ are identical:

$$\mathrm{VaR}_\alpha(X) = q_\alpha(X) = \min\{z \,|\, F_X(z) \geq \alpha\}. \tag{5.1}$$

The *conditional value-at-risk* $\mathrm{CVaR}_\alpha(X)$ is defined by

$$\mathrm{CVaR}_\alpha(X) = \text{expectation of } X \text{ in the conditional distribution of its upper } \alpha\text{-tail}, \tag{5.2}$$

so that, in particular,

$$\mathrm{CVaR}_\alpha(X) \geq \mathrm{VaR}_\alpha(X) \quad \text{always}. \tag{5.3}$$

The specification of what is meant by the "upper α-tail" requires careful examination to clear up ambiguities. It should refer to the outcomes $X(\omega)$ in the upper part of the range of X for which the probability is $1-\alpha$. Ordinarily this would be the interval $[q_\alpha(X), \infty)$, but that is not possible when there is a probability atom of size $\delta > 0$ at $q_\alpha(X)$ itself (corresponding to F_X having a jump of such size at $q_\alpha(X)$), because then $\mathrm{prob}[q_\alpha(X), \infty)$ is not necessarily $1-\alpha$ but rather something between $\mathrm{prob}(q_\alpha(X), \infty)$ and $\mathrm{prob}(q_\alpha(X), \infty) + \delta$. The α-tail conditional distribution cannot then be just the restriction of the P distribution to the interval $[q_\alpha(X), \infty)$, rescaled by dividing by $1-\alpha$. Instead, that rescaling has to be applied to the distribution obtained on $[q_\alpha(X), \infty)$ obtained by splitting the probability atom at $q_\alpha(X)$ so as to leave just enough to bring the total probability up to $1-\alpha$.

Although this is a foolproof definition for clarifying the concept, as introduced in Rockafellar and Uryasev [14] as a follow-up to the original definition of CVaR in Rockafellar and Uryasev [13], other formulas for $\mathrm{CVaR}_\alpha(X)$ may be operationally more convenient in some situations. One of these is

$$\mathrm{CVaR}_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{VaR}_\beta(X)\, d\beta, \tag{5.4}$$

coming from Acerbi, cf. [1]. It has led to the term *average value-at-risk* being preferred for this concept in Föllmer and Schied [7], which has become a major reference in financial mathematics, although Acerbi preferred *expected shortfall* (a term with an orientation that conflicts with our cost view of random variables X). The most potent formula for CVaR may be the *minimization rule*

$$\mathrm{CVaR}_\alpha(X) = \min_{C \in \mathbb{R}} \{C + (1-\alpha)^{-1} E[\max\{0, X - C\}]\}, \tag{5.5}$$

which was established in Rockafellar and Uryasev [13, 14]. It has the illuminating counterpart that

$$\mathrm{VaR}_\alpha(X) = \text{left endpoint of } \operatorname*{arg\,min}_{C \in \mathbb{R}} \{C + (1-\alpha)^{-1} E[\max\{0, X - C\}]\}. \tag{5.6}$$

In (5.5) and (5.6) a continuous convex function of $C \in \mathbb{R}$ (dependent on X and α) is being minimized over $\mathbb{R}$. The arg min giving the values of C for which the minimum is attained is, in this special case, known always to be a nonempty, closed, bounded interval. Much of the time that interval collapses to a single value, $\mathrm{VaR}_\alpha(X)$, but if not, then $\mathrm{VaR}_\alpha(X)$ is the lowest value in it.

We have posed in this formula in terms of $\mathrm{VaR}_\alpha(X)$ for harmony with $\mathrm{CVaR}_\alpha(X)$, but it can easily be seen as an expression for calculating $q_\alpha(X)$. In that respect, it is equivalent to a formula of Koenker and Bassett [9], which is central to quantile regression (Koenker [8])

$$q_\alpha(X) = \text{left endpoint of } \operatorname*{arg\,min}_{C \in \mathbb{R}} E[\max\{0, X - C\} + (\alpha^{-1} - 1)\max\{0, C - X\}]. \tag{5.7}$$

Researchers in that area paid little attention to the minimum value in (5.5), but that value is primary here for the following reason.

Theorem 2. *For any probability level $\alpha \in (0,1)$, the functional $\mathcal{R}(X) = \mathrm{CVaR}_\alpha(X)$ is a coherent measure of risk in the basic sense.*

This conclusion was reached from several directions. After the concept of conditional value-at-risk was introduced in Rockafellar and Uryasev [13] along with the minimization rule in (5.5), but initially only under the simplifying assumption that F_X had no jumps, Pflug [10] proved that a functional given by the right side of (5.5) would be a coherent measure of risk even if jumps were present. The fact that $\mathrm{CVaR}_\alpha(X)$, if extended to the case of jumps by the careful interpretation of the α-tail in (5.2), would still hold (5.5) was proved in Rockafellar and Uryasev [14]. Meanwhile, Acerbi and Tasche [2] showed the coherency of functionals expressed by the right side of (5.4), thus covering CVaR in another way.

Prior to these efforts, even with the strong case for coherent risk measures originally made in Artzner et al. [3, 4] essentially *no* good example of such a measure had been identified that had practical import beyond the theoretical. (In Artzner et al. [4], only general examples corresponding to the risk envelope formula to be presented in Theorem 4(a) were provided, without specifics. Other proposed measures such as tail risk, bearing a resemblance to conditional value-at-risk, but neglecting the complication of probability atoms, were, for a while, believed to offer coherence.) Currently, conditional value-at-risk, and other measures of risk constructed from it, form the core of the subject, especially in view of deep connections found in utility theory and second-order stochastic dominance.

A further property of conditional value-at-risk, distinguishing it from value-at-risk, is that for any single $X \in \mathcal{L}^2$,

$$\begin{aligned} &\mathrm{CVaR}_\alpha(X) \text{ depends continuously on } \alpha \in (0,1), \\ &\text{with } \lim_{\alpha\to 1} \mathrm{CVaR}_\alpha(X) = \sup X \text{ and } \lim_{\alpha\to 0} \mathrm{CVaR}_\alpha(X) = EX. \end{aligned} \tag{5.8}$$

For value-at-risk, the limits are the same, but the dependence is not always continuous.

To appreciate the implications of conditional value-at-risk in approaching uncertainty, it will help to look first at what happens with value-at-risk itself. The crucial observation is that, through the definition in (5.1), one has

$$\mathrm{prob}\{X \le 0\} \ge \alpha \iff q_\alpha(X) \le 0 \iff \mathrm{VaR}_\alpha(X) \le 0. \tag{5.9}$$

This can be used to rewrite the probabilistic constraints of Approach 5 in (2.6) and also the associated objective in (2.7), since

$$\mathrm{prob}\{X \le c\} \ge \alpha \iff q_\alpha \le c \iff \mathrm{VaR}_\alpha(X) \le c. \tag{5.10}$$

In this way, Approach 5 can be expressed in the same kind of pattern as the others, where the random variables $\underline{c}_i(x)$ are composed with some $\mathcal{R}_i$ as in problem (3.1).

5.1. Approach 5, Recast: Safeguarding with Value-at-Risk

For a choice of probability levels $\alpha_i \in (0,1)$ for $i = 0, 1, \ldots, m$,

$$\begin{aligned} &\text{minimize } \mathrm{VaR}_{\alpha_0}(\underline{c}_0(x)) \quad \text{over all } x \in S \text{ satisfying} \\ &\mathrm{VaR}_{\alpha_i}(\underline{c}_i(x)) \le 0 \text{ for } i = 1, \ldots, m. \end{aligned} \tag{5.11}$$

In this case we have $\mathcal{R}_i(X) = \mathrm{VaR}_{\alpha_i}(X) = q_{\alpha_i}(X)$ and the original interpretations of this approach hold: We are asking in the constraints that the outcome of $\underline{c}_i(x)$ should lie in $(-\infty, 0]$ with probability at least α_i, and subject to that, are seeking the lowest value c such that $\underline{c}_0(x) \le c$ with probability at least α_0. But the shortcomings still hold as well: $\mathcal{R}_i$ is *not* a coherent measure of risk, even though it satisfies (R1), (R3), (R4), and (R5). In looking to it for guidance, one could be advised paradoxically against diversification.

From a technical standpoint, other poor features of using $\mathrm{VaR}_\alpha(X)$, or in equivalent notation $q_\alpha(X)$, to assess the overall risk in a random variable X are revealed. The formula in (5.1) predicts discontinuous behavior when dealing with random variables X whose distri-

bution functions F_X may have graphical flat spots or jumps, as is inevitable with discrete, and in particular, empirical distributions. All this is avoided by working with conditional value-at-risk instead. Those familiar with the fundamentals of optimization can immediately detect the root of the difference by comparing the formulas in (5.5) in (5.6). It is well known that the minimum value in an optimization problem dependent on parameters, even a problem in one dimension, as in this case, behaves much better than does the solution, or set of solution points. Thus, $\mathrm{CVaR}_\alpha(X)$ should behave better as a function of X than $\mathrm{VaR}_\alpha(X)$ and indeed it does.

5.2. Approach 6: Safeguarding with Conditional Value-at-Risk

For a choice of probability levels $\alpha_i \in (0,1)$ for $i = 0, 1, \ldots, m$,

$$\begin{aligned} &\text{minimize } \mathrm{CVaR}_{\alpha_0}(\underline{c}_0(x)) \text{ over all } x \in S \text{ satisfying} \\ &\mathrm{CVaR}_{\alpha_i}(\underline{c}_i(x)) \le 0 \text{ for } i = 1, \ldots, m. \end{aligned} \tag{5.12}$$

Here we use the coherent risk measures $\mathcal{R}_i = \mathrm{CVaR}_{\alpha_i}$. What effect does this have on the interpretation of the model, in contrast to that of Approach 5, where $\mathcal{R}_i = \mathrm{VaR}_{\alpha_i}$? The conditional expectation in the definition of conditional value-at-risk provides the answer. However, due to the small complications that can arise over the meaning of the upper α_i-tail of the random variable $\underline{c}_i(x)$ when its distribution function may have a jump at the quantile value $q_{\alpha_i}(\underline{c}_i(x)) = \mathrm{VaR}_{\alpha_i}(\underline{c}_i(x))$, it is best, for an initial understanding of the idea, to suppose there are no such jumps. Then,

$$\begin{aligned} &\mathrm{CVaR}_{\alpha_i}(\underline{c}_i(x)) \le 0 \text{ means not merely that } \underline{c}_i(x) \le 0 \\ &\quad \text{at least } 100\alpha_i\% \text{ of the time, but that the average of the} \\ &\quad \text{worst } 100(1-\alpha_i)\% \text{ of all possible outcomes will be } \le 0. \end{aligned} \tag{5.13}$$

Obviously, Approach 6 is, in this way, more cautious than Approach 5.

A valuable feature in Approach 6 is the availability of the minimization rule (5.5) for help in solving a problem in formulation (5.12). Insert this formula, with an additional optimization variable C_i for each index i, and the resultant problem to solve is to

$$\begin{aligned} &\text{find } (x, C_0, C_1, \ldots, C_m) \in X \times \mathbb{R}^{m+1} \text{ to minimize} \\ &\quad C_0 + (1-\alpha_0)^{-1} E[\max\{0, \underline{c}_0(x) - C_0\}] \text{ subject to} \\ &\quad C_i + (1-\alpha_i)^{-1} E[\max\{0, \underline{c}_i(x) - C_i\}] \le 0, \quad i = 1, \ldots, m. \end{aligned} \tag{5.14}$$

Especially interesting is the case where each $c_i(x, \omega)$ depends linearly (affinely) on x, and the space Ω of future states ω is finite. The expectations become weighted sums in which, through the introduction of still more variables, each max term can be replaced by a pair of linear inequalities so as to arrive at a *linear programming* reformulation of (5.14); cf. Rockafellar and Uryasev [14].

5.3. Staircasing

As a reminder, these approaches are being described in the direct picture of problem (3.1), but they also encompass the finer possibilities associated with breaking a random variable down into a staircased sequence, as in (2.11), obtained from a series of cost thresholds. (See comment after (3.1).)

6. Further Examples and Risk Envelope Duality

The examples of coherent measures of risk that we have accumulated so far are in Approaches 1, 2, 3, and 6. Other prime examples are provided in this section, including some that fit only the extended, not the basic, definition of coherency. Note, however, that any collection of examples automatically generates an even larger collection through the following operations, as is seen from the definition of coherency.

Theorem 3. *Coherency-preserving operations.*

(a) *If $\mathcal{R}_1, \ldots, \mathcal{R}_r$ are coherent measures of risk in the basic sense, and if $\lambda_1, \ldots, \lambda_r$ are positive coefficients adding to* 1*, then a coherent measure of risk is defined by*

$$\mathcal{R}(X) = \lambda_1 \mathcal{R}_1(X) + \lambda_2 \mathcal{R}_2(X) + \cdots + \lambda_r \mathcal{R}_r(X). \tag{6.1}$$

Moreover, the same holds for coherent measures of risk in the extended sense.

(b) *If $\mathcal{R}_1, \ldots, \mathcal{R}_r$ are coherent measures of of risk in the basic sense, then so too is*

$$\mathcal{R}(X) = \max\{\mathcal{R}_1(X), \mathcal{R}_2(X), \ldots, \mathcal{R}_r(X)\}. \tag{6.2}$$

Moreover, the same holds for coherent measures of risk in the extended sense.

6.1. Mixed CVaR and Spectral Profiles of Risk

An especially interesting example of such operations is *mixed conditional value-at-risk*, which refers to functionals having the form

$$\mathcal{R}(X) = \lambda_1 \mathrm{CVaR}_{\alpha_1}(X) + \cdots + \lambda_r \mathrm{CVaR}_{\alpha_r}(X) \quad \text{with } \alpha_i \in (0,1), \lambda_i > 0, \lambda_1 + \cdots + \lambda_r = 1. \tag{6.3}$$

These functionals likewise furnish coherent measures of risk in the basic sense. One can even extend this formula to continuous sums

$$\mathcal{R}(X) = \int_0^1 \mathrm{CVaR}_\alpha(X)\, d\lambda(\alpha). \tag{6.4}$$

This gives a coherent measure of risk in the basic sense for any weighting measure λ (with respect to generalized integration) that is nonnegative with total weight equal to 1. The formula in (6.3) corresponds to the discrete version of (6.4) in which a probability atom of size λ_i is placed at each α_i. It has been proved (see Rockafellar et al. [16, Proposition 5]) that as long as $\int_0^1 (1-\alpha)^{-1}\, d\lambda(\alpha) < \infty$, the measure of risk in (6.4) has an alternative expression in the form

$$\mathcal{R}(X) = \int_0^1 \mathrm{VaR}_\alpha(X)\phi(\alpha)\, d\alpha \quad \text{with } \phi(\alpha) = \int_{(0,\alpha]} (1-\beta)^{-1}\, d\lambda(\beta). \tag{6.5}$$

This is a *spectral representation*, in the sense of Acerbi [1], which relates to a dual theory of utility where ϕ gives the *risk profile* of the decision maker.

Clearly, risk measures of form (6.3) can be used to approximate risk measures like (6.4). On the other hand, such risk measures can be supplied through the minimization rule (5.5) with a representation in terms of parameters $C_1, \ldots, C_r$ which is conducive to their practical use in optimization along the lines of the prescription at the end of the preceding section.

6.2. Approach 7: Safeguarding with Mixtures of Conditional Value-at-Risk

The use of single CVaR risk measures in Approach 6 could be expanded to mixtures as just described, with possible connections to risk profiles. All the properties in Theorem 1 would be available.

6.3. Risk Measures from Subdividing the Future

Let Ω be partitioned into subsets $\Omega_1, \ldots, \Omega_r$ having positive probability, and for $k = 1, \ldots, r$ let

$$\mathcal{R}_k(X) = \sup_{\omega \in \Omega_k} X(\omega). \tag{6.6}$$

Then R_k is a coherent measure of risk in the basic sense (just like the one in Approach 2), and so too then, by Theorem 3(a), is

$$\mathcal{R}(X) = \lambda_1 \sup_{\omega\in\Omega_1} X(\omega) + \cdots + \lambda_r \sup_{\omega\in\Omega_r} X(\omega) \quad \text{for coefficients } \lambda_i > 0 \text{ adding to 1.} \tag{6.7}$$

The weights λ_k could be seen as lending different degrees of importance to different parts of Ω. They could also be viewed as providing a sort of skeleton of probabilities to Ω. A better understanding of this will be available below; see (6.16) and the surrounding explanation.

6.4. Approach 8: Distributed Worst-Case Analysis

This refers to the modification of the worst-case formulation in Approach 2 to encompass risk measures of the forms in (6.6) and (6.7). Different partitions of Ω might be used for different constraints.

6.5. Risk Measures of Penalty Type

Another interesting way of quantifying the risk of loss is to modify the expected cost by adding a penalty term for positive costs. Recall that the so-called $\mathcal{L}^p$*-norms* are well defined as functionals on $\mathcal{L}^2$ by

$$||X||_p = \begin{cases} E|X| & \text{for } p = 1, \\ (E[\,|X|^p])^{1/p} & \text{for } 1 < p < \infty, \\ \sup|X| & \text{for } p = \infty, \end{cases} \tag{6.8}$$

with $\|X\|_p \le \|X\|_{p'}$ when $p \le p'$, but $\|X\|_p$ can take on ∞ when $p > 2$ and Ω is not finite (in which case $\|\cdot\|_p$ is not technically a "norm" any more on $\mathcal{L}^2$). Consider the functional $\mathcal{R}\colon \mathcal{L}^2 \to (-\infty, \infty]$ defined by

$$\mathcal{R}(X) = EX + \lambda||\max\{0, X - EX\}||_p \quad \text{with } p \in [1,\infty],\ \lambda \in [0,1]. \tag{6.9}$$

This too gives a coherent measure of risk in the basic sense. The coherency in (6.9) is not hard to verify up to a point: Axioms (R1), (R2), and (R4) are easily checked, along with (R5) for scalability. The monotonicity in (R3), however, is a bit more daunting. It is seen through the equivalence of (R3) with (3.6) using the inequality that

$$\|\max\{0, X - EX\}\|_p \le \|\max\{0, X - EX\}\|_\infty = \sup X - EX$$

and the observation that $EX + \lambda(\sup X - EX) \le 0$ when $X \le 0$ and $0 \le \lambda \le 1$.

Often in financial applications where $c_0(x,\omega)$ refers to the shortfall relative to a specified target level of profit, a penalty expression is like $||\max\{0, \underline{c}_0(x)\}||_p$ is minimized, or such an expression raised to a power $a > 1$. This corresponds to composing $\underline{c}_0(x)$ with $\mathcal{R}(X) = ||\max\{0, X\}||_p^a$, which is *not* a coherent measure of risk. It satisfies (R2), (R3), (R4), and when $a = 1$ even (R5). Only (R1) fails. The convexity preservation in Theorem 1(a) would hold, although not the certainty preservation in Theorem 1(b). A shortcoming is in the absence of a control over the expected value of $\underline{c}_0(x)$, which might even be positive. Minimum penalty might be achieved by a decision x in which there is a very high probability of loss, albeit not a big loss. By contrast, however, composition with a coherent risk measure such as in (6.9) would facilitate creating a safety margin against loss.

6.6. Log-Exponential Risk Measures

Until now, every coherent measure of risk has satisfied (R5). Here is an important one that does not and therefore must be considered in the extended sense rather than the basic sense of coherency

$$\mathcal{R}(X) = \lambda \log E[e^{X/\lambda}] \text{ for a parameter value } \lambda > 0. \tag{6.10}$$

The confirmation of coherency in this case will be based on the duality theory presented next.

6.7. Representations of Risk Measures by Risk Envelopes

Coherent measures of risk can always be interpreted as coming from a kind of augmented worst-case analysis of expectations with respect to other probability measures P' on Ω than the nominal one, P, or more specifically such measures P' having a well defined density $Q = dP'/dP \in \mathcal{L}^2$ with respect to P. Such densities functions make up the set

$$\mathcal{P} = \{Q \in \mathcal{L}^2 \mid Q \geq 0,\ EQ = 1\}. \tag{6.11}$$

When Q is the density for P', the expectation of a random variable X with respect to P' instead of P is $E[XQ]$, inasmuch as

$$E[XQ] = \int_\Omega X(\omega)Q(\omega)\,dP(\omega) = \int_\Omega X(\omega)\frac{dP'}{dP}(\omega)\,dP(\omega) = \int_\Omega X(\omega)\,dP'(\omega). \tag{6.12}$$

In this framework P itself corresponds to $Q \equiv 1$: We have $EX = E[X \cdot 1]$.

In contemplating a subset $\mathcal{Q}$ of $\mathcal{P}$, one is essentially looking at some collection of alternatives P' to P. This could be motivated by a reluctance to accept P as furnishing a completely reliable model for the relative occurrences of the future states $\omega \in \Omega$, and the desire to test the dangers of too much trust in P.

A *risk envelope* will mean a nonempty, convex subset $\mathcal{Q}$ of $\mathcal{P}$ that is closed (so when elements Q^k of $\mathcal{Q}$ converge to some Q in the $\mathcal{L}^2$ sense described earlier, Q also belongs to $\mathcal{Q}$). An *augmented risk envelope* will mean a risk envelope $\mathcal{Q}$ supplied with a function $a\colon \mathcal{Q} \to [0, \infty]$ (the *augmenting function*) having the properties that

$$\begin{cases} \text{the set } \{Q \in \mathcal{Q} \mid a(Q) < \infty\} \text{ has } \mathcal{Q} \text{ as its closure,} \\ \text{the set } \{Q \in \mathcal{Q} \mid a(Q) \leq C\} \text{ is closed for } C < \infty, \\ \text{the function } a \text{ is convex on } \mathcal{Q} \text{ with } \inf\limits_{Q \in \mathcal{Q}} a(Q) = 0. \end{cases} \tag{6.13}$$

As a special case one could have $a \equiv 0$ on $\mathcal{Q}$, and then the idea of an augmented risk envelope would simply reduce to that of a risk envelope by itself.

Theorem 4. *Dual characterization of coherency.*

(a) *$\mathcal{R}$ is a coherent measure of risk in the basic sense if and only if there is a risk envelope $\mathcal{Q}$ (which will be uniquely determined) such that*

$$\mathcal{R}(X) = \sup_{Q \in \mathcal{Q}} E[XQ]. \tag{6.14}$$

(b) *$\mathcal{R}$ is a coherent measure of risk in the extended sense if and only if there is a risk envelope $\mathcal{Q}$ with an augmenting function a (both of which will be uniquely determined) such that*

$$\mathcal{R}(X) = \sup_{Q \in \mathcal{Q}} \{E[XQ] - a(Q)\}. \tag{6.15}$$

The proof of this key result, reflecting a basic conjugacy principle in convex analysis (Rockafellar [11, 12]), can be found in a number of places, subject to variations on the underlying space (not always $\mathcal{L}^2$). A version for the scalable case with Ω finite appeared in Artzner et al. [4] and was elaborated for infinite Ω in the unpublished exposition of Delbaen [6]. It was taken up specially for $\mathcal{L}^2$ in Rockafellar et al. [17] where the term risk envelope was introduced. (The main results of that working paper were eventually published in Rockafellar et al. [16].) Versions without scalability are in Föllmer and Schied [7] and Ruszczyński and Shapiro [20]. The condition in (6.13), that $\inf_{\mathcal{Q}} a = 0$, is essential for getting axiom (R1) to be satisfied in (6.15).

In view of the preceding discussion, formula (6.14) for the basic case has the interpretation that the risk $\mathcal{R}(X)$ in X comes simply from a *worst-case analysis of the expected costs $E[XQ]$* corresponding to the probability measures P' having densities Q in the specified set $\mathcal{Q}$.

In short, selecting a coherent risk measure $\mathcal{R}$ is equivalent to selecting a risk envelope $\mathcal{Q}$. Of course, this is rather black and white. Either a density Q presents a concern or it does not. The extended case with an augmenting function a provides a gradation in (6.15): Densities Q have less influence when $a(Q) > 0$.

Formula (6.14) would still give a coherent risk measure $\mathcal{R}$ with $\mathcal{Q}$ taken to be *any* nonempty subset $\mathcal{Q}_0$ of $\mathcal{P}$, but that $\mathcal{R}$ would also then be given by taking $\mathcal{Q}$ to be the closed convex hull of $\mathcal{Q}_0$ (the smallest closed convex subset of $\mathcal{L}^2$ that includes $\mathcal{Q}$). The assumption that $\mathcal{Q}$ is already closed and convex makes it possible to claim that (6.12) furnishes a one-to-one correspondence $\mathcal{R} \leftrightarrow \mathcal{Q}$. A similar statement applies to formula (6.15).

6.8. Risk Envelope for Guessing the Future

The coherent (but hardly to be recommended) risk measure $\mathcal{R}(X) = X(\bar{\omega})$ for a future state $\bar{\omega}$ with $\text{prob}(\bar{\omega}) > 0$ corresponds to taking $\mathcal{Q}$ in (6.14) to consist of a single function Q, which has $Q(\bar{\omega}) = 1/\text{prob}(\bar{\omega})$ but $Q(\omega) = 0$ otherwise.

6.9. Risk Envelope for Worst-Case Analysis

The coherent risk measure $\mathcal{R}(X) = \sup X$ corresponds to taking $\mathcal{Q}$ in (6.14) to be all of $\mathcal{P}$, i.e., to consist of all $Q \geq 0$ with $EQ = 1$.

6.10. Risk Envelope for Distributed Worst-Case Analysis

In the broader setting of Ω being partitioned into subsets Ω_k with weights λ_k as in (6.7), the risk envelope $\mathcal{Q}$ consists of the densities Q with respect to P of the probability measures P' such that

$$P'(\Omega_k) = \lambda_k \quad \text{for } k = 1, \ldots, r. \tag{6.16}$$

Not all probability measures alternative to P are admitted, as with ordinary worst-case analysis, but only those that conform to a specified framework of the likelihoods of different parts of Ω. This provides a means for incorporating a rough structure of probabilities without having to go all the way to a particular measure like P, which serves here only in the technical background.

6.11. Risk Envelope for Relying on Expectations

The coherent risk measure for $\mathcal{R}(X) = \mu(X)$ corresponds to taking $\mathcal{Q}$ in (6.14) to consist solely of $Q \equiv 1$.

6.12. Risk Envelope for Standard Deviation Units as Safety Margins?

For the functional $\mathcal{R}(X) = \mu(X) + \lambda\sigma(X)$ there is no risk envelope $\mathcal{Q} \subset \mathcal{P}$, due to the absence of coherency. However, because only (R3) fails, there is a representation in the form (6.14) involving elements Q that are not necessarily ≥ 0. (See Rockafellar et al. [16].)

6.13. Risk Envelope for Safeguarding with Value-at-Risk, or in Other Words, for Specifying Probabilities of Compliance?

For $\mathcal{R}(X) = q_\alpha(X) = \text{VaR}_\alpha(X)$ there is no corresponding risk envelope $\mathcal{Q}$, and in fact no representation in the pattern of (6.14), because $\mathcal{R}$ lacks convexity.

6.14. Risk Envelope for Safeguarding with Conditional Value-at-Risk

For the functional $\mathcal{R}(X) = \text{CVaR}_\alpha(X)$, the risk envelope is

$$\mathcal{Q} = \{Q \in \mathcal{P} \mid Q \leq 1/\alpha\}. \tag{6.17}$$

This was first shown in Rockafellar et al. [17]. (See also Rockafellar et al. [16].)

6.15. Risk Envelope for Mixed Conditional Value-at-Risk

For $\mathcal{R}(X) = \sum_{i=1}^{r} \lambda_i \mathrm{CVaR}_{\alpha_i}(X)$ with positive weights λ_i adding to 1, the risk envelope is

$$\mathcal{Q} = \left\{ \sum_{i=1}^{r} \lambda_i Q_i \,\middle|\, Q_i \in \mathcal{P},\ 0 \le Q_i \le 1/\alpha_i \right\}. \tag{6.18}$$

Again, this comes from Rockafellar et al. [17]. (See also Rockafellar et al. [16].)

6.16. Risk Envelope for Measures of Penalty Type

For $\mathcal{R}(X) = EX + \lambda \| \max\{0, X - EX\} \|_p$ with $\lambda > 0$ and $p \in [1, \infty]$, the risk envelope is

$$\mathcal{Q} = \left\{ Q \in \mathcal{P} \,\middle|\, \|Q - \inf Q\|_q \le 1 \right\} \quad \text{where } q = \begin{cases} (1 - p^{-1})^{-1} & \text{when } p < \infty, \\ 1 & \text{when } p = \infty. \end{cases} \tag{6.19}$$

The proof of this is found in Rockafellar et al. [16, Examples 8 and 9]. (In all such references to the literature, the switch of orientation from X giving rewards to X giving costs requires a switch in signs.)

6.17. Augmented Risk Envelope for Log-Exponential Risk

The measure of risk expressed by $\mathcal{R}(X) = \lambda \log E[e^{X/\lambda}]$, which is not positively homogeneous, requires a risk envelope $\mathcal{Q}$ with an augmenting function a, namely

$$\mathcal{Q} = \mathcal{P} \quad \text{with } a(Q) = \lambda E[Q \log Q]. \tag{6.20}$$

This recalls the duality between log-exponential functions and entropy-like functions which is well known in convex analysis, cf. Rockafellar [11]; Rockafellar and Wets [15, p. 482]. Its application to risk measures can be seen in Föllmer and Schied [7, p. 174], who refer to this $\mathcal{R}$ as an "entropy" risk measure. The coherency of comes from showing that $\mathcal{R}$ is obtained from the $\mathcal{Q}$ and a in (6.20) by the formula in Theorem 4(b). In terms of the probability measure P' having density $Q = dP'/dP$, one has

$$E[Q \log Q] = I(P', P), \quad \text{the relative entropy of } P' \text{ over } P. \tag{6.21}$$

Ben-Tal and Teboulle [5] open this further and note that $E[a(Q)]$ has been studied for more general a than $a(Q) = [Q \log Q]$, as in (6.20), for which they supply references.

7. Safety Margins and Measures of Deviation

Although the safety margins in Approach 4, using units of standard deviation, collide with coherency, the concept of a safety margin is too valuable to be ignored. The key idea behind it is to remedy the weakness of an expected cost constraint $E[\underline{c}_i(x)] \le 0$ by insisting on an adequate barrier between $E[\underline{c}_i(x)]$ and 0. Is this ensured simply by passing to a constraint model $\mathcal{R}_i(\underline{c}_i(x)) \le 0$ for a coherent measure of risk $\mathcal{R}_i$, as we have been considering? Not necessarily. Guessing the future, with $\mathcal{R}_i(X) = X(\bar{\omega})$ for some $\bar{\omega}$ of positive probability, provides a quick counterexample. There is no reason to suppose that having $X(\bar{\omega}) \le 0$ entails having $EX \le 0$. We have to impose some restriction on $\mathcal{R}_i$ to get a safety margin. The following class of functionals must be brought in.

7.1. Averse Measures of Risk

Relative to the underlying probability measure P on Ω, a functional $\mathcal{R}\colon \mathcal{L}^2 \to (-\infty, \infty]$ will be called an *averse measure of risk in the extended sense*, if it satisfies axioms (R1), (R2), (R4) and

(R6) $\mathcal{R}(X) > EX$ for all nonconstant X,

and *in the basic sense*, if it also satisfies (R5).

Recall that (R1) guarantees $\mathcal{R}(X) = EX$ for constant $X \equiv C$. Aversity has the interpretation that *the risk of loss in a nonconstant random variable X cannot be acceptable unless, in particular, $X(\omega) < 0$ on average*. Note that relations to expectation, and consequently to the particular choice of P, have not entered axiomatically until this point. (Averse measures of risk were initially introduced in Rockafellar et al. [17] in terms of "strict expectation-boundedness" rather than "aversity." See also Rockafellar et al. [16].)

The monotonicity in (R3) has not been required in the definition, so an averse measure of risk might not be coherent. On the other hand, a coherent measure might not be averse, as the preceding illustration makes clear. In the end, we will want to focus on measures of risk that are simultaneously averse relative to P and coherent. At this stage, however, the concepts will come out clearer if we do not insist on that.

Averse measures of risk relative to P will be crucial in making the connection with the other fundamental way of quantifying the uncertainty in a random variable, namely its degree of deviation from constancy. Next, we develop this other kind of quantification axiomatically.

7.2. Measures of Deviation

A functional $\mathcal{D}\colon \mathcal{L}^2 \to [0, \infty]$ will be called a *measure of deviation in the extended sense* if it satisfies

(D1) $\mathcal{D}(C) = 0$ for constants C, but $\mathcal{D}(X) > 0$ for nonconstant X,
(D2) $\mathcal{D}((1-\lambda)X + \lambda X') \le (1-\lambda)\mathcal{D}(X) + \lambda\mathcal{D}(X')$ for $\lambda \in (0,1)$ ("convexity").
(D3) $\mathcal{D}(X) \le d$ when $||X^k - X||_2 \to 0$ with $\mathcal{D}(X^k) \le d$ ("closedness").

It will be called a *measure of deviation in the basic sense* when furthermore

(D4) $\mathcal{D}(\lambda X) = \lambda\mathcal{D}(X)$ for $\lambda > 0$ ("positive homogeneity").

Either way, it will be called *coherent* if it also satisfies

(D5) $\mathcal{D}(X) \le \sup X - EX$ for all X ("upper range boundedness")

An immediate example of a measure of deviation in the basic sense is standard deviation, $\mathcal{D}(X) = \sigma(X)$. In addition to (D1), (D2), and (D3), it satisfies (D4), but *not*, as it turns out, (D5). The definition aims to quantify the uncertainty, or nonconstancy, of X in different ways than just standard deviation, allowing even for cases where $\mathcal{D}(X)$ might not be the same as $\mathcal{D}(-X)$. The reason for tying (D5) to $\mathcal{D}$ being "coherent" is revealed by the theorem below.

7.3. Risk Measures Versus Deviation Measures

Theorem 5. *A one-to-one correspondence between measures of deviation $\mathcal{D}$ in the extended sense and averse measures of risk $\mathcal{R}$ in the extended sense is expressed by the relations*

$$\mathcal{R}(X) = EX + \mathcal{D}(X), \qquad \mathcal{D}(X) = \mathcal{R}(X - EX), \tag{7.1}$$

with respect to which

$$\mathcal{R} \text{ is coherent} \iff \mathcal{D} \text{ is coherent.} \tag{7.2}$$

In this correspondence, measures in the basic sense are preserved:

$$\mathcal{R} \text{ is positively homogeneous} \iff \mathcal{D} \text{ is positively homogeneous.} \tag{7.3}$$

This result, for the basic case, was originally obtained in the working paper Rockafellar et al. [17] and finally in Rockafellar et al. [16]. Extension of the proof beyond positive homogeneity is elementary. The risk envelope representation of Theorem 4 for $\mathcal{R}$ under coherency immediately translates, when $\mathcal{R}$ is strict, to a similar representation for the associated $\mathcal{D}$

$$\mathcal{D}(X) = \sup_{Q\in\mathcal{Q}} E[(X - EX)Q] \quad \text{in the basic case} \tag{7.4}$$

or, on the other hand

$$\mathcal{D}(X) = \sup_{Q\in\mathcal{Q}} \{E[(X - EX)Q] - E[a(Q)]\} \quad \text{in the extended case.} \tag{7.5}$$

It follows from Theorem 5 that a deviation measure $\mathcal{D}$ is coherent *if and only if* it has a representation of this kind (necessarily unique) in which $\mathcal{Q}$ and a (identically 0 in the basic case) meet the specifications in (6.13). Deviation measures that are not coherent have representations along the same lines, but with $\mathcal{Q} \not\subset \mathcal{P}$. The elements $Q \in \mathcal{Q}$ still have $EQ = 1$, but $Q \not\geq 0$ for some, and their interpretation in terms of densities dP'/dP of probability measures P' being compared to P drops away. For more on this, see Rockafellar et al. [16].

The fact that standard deviation $\mathcal{D}(X) = \sigma(X)$ does not have a risk envelope representation (7.4) with $\mathcal{Q} \subset \mathcal{P}$ lies behind the assertion that this deviation measure is not coherent and, at the same time, confirms the lack of that property in Approach 4. This shortcoming of $\mathcal{D}(X) = \sigma(X)$ can also be gleaned from the condition for $\mathcal{D}$ to be coherent in (D5). If $\sigma(X) \leq \sup X - EX$ for all X, we would also have by applying this to $-X$, that $\sigma(X) \leq EX - \inf X$, and therefore $\sigma(X) \leq [\sup X - \inf X]/2$ for all random variables X, which is in general false.

Theorem 5 shows that to introduce safety margins in optimization under uncertainty without falling into the trap of Approach 4 with its lack of coherency, we must pass from standard deviation units to those of some other measure of deviation satisfying (D5). Here we can take advantage of the fact that when $\mathcal{D}$ is a coherent measure of deviation, then so too is $\lambda\mathcal{D}$ for any $\lambda > 0$.

7.4. Approach 9: Generalized Deviation Units as Safety Margins

Faced with the random variables (1.2), choose deviation measures $\mathcal{D}_i$ for $i = 0, 1, \ldots, m$ with coefficients $\lambda_i > 0$. Pose the constraints in the form

$$\mu(\underline{c}_i(x)) + \lambda_i \mathcal{D}_i(\underline{c}_i(x)) \leq 0 \quad \text{for } i = 1, \ldots, m, \tag{7.6}$$

thus requiring that positive outcomes of $\underline{c}_i(x)$ can only occur in the part of the range of this random variable lying more than λ_i deviation units, with respect to $\mathcal{D}_i$, above the mean $\mu(\underline{c}_i(x)) = E[\underline{c}_i(x)]$. The goal is to

$$\begin{aligned} &\text{minimize } \mu(\underline{c}_0(x) - x_{n+1}) + \lambda_0 \mathcal{D}_i(\underline{c}_0(x) - x_{n+1}) \\ &\qquad \text{over all } (x, x_{n+1}) \in S \times \mathbb{R} \text{ satisfying (7.4).} \end{aligned} \tag{7.7}$$

This mimics Approach 4 with $\sigma(X)$ replaced by $\mathcal{D}_i(X)$. The measures of risk filling the role prescribed in (3.1) now have the form

$$\mathcal{R}_i(X) = \mu(X) + \lambda_i \mathcal{D}_i(X) = EX + \mathcal{D}'_i(X) \quad \text{for } i = 1, \ldots, m, \text{ with } \mathcal{D}'_i = \lambda_i \mathcal{D}_i. \tag{7.8}$$

The contrast between this and the previous case, where $\mathcal{D}_i(X) = \sigma(X)$, is that $\mathcal{R}_i$ *is coherent when* $\mathcal{D}_i$ *is coherent*. Hence, if this holds for $i = 0, 1, \ldots, m$, the optimization properties in Theorem 1 apply to problem (7.7).

Theorem 5 provides the right that all the approaches to optimization under uncertainty considered so far in the mode of (3.1) in which the $\mathcal{R}_i$s are *averse* measures of risk correspond

to introducing safety margins in units of generalized deviation. But this is not true when the $\mathcal{R}_i$s are not averse. Because we want the $\mathcal{R}_i$s to be coherent as well, the question arises: What examples do we have at this point of *coherent* measures of risk that are also *averse* (relative to P)? Those obtained via (7.8) from a coherent measure of deviation serve the purpose, but the issue then devolves to looking for examples of such measures of deviation.

7.5. Aversity of CVaR and Mixed CVaR

The coherent risk measure $\mathcal{R}(X) = \mathrm{CVaR}_\alpha(X)$ is averse for any $\alpha \in (0,1)$. The corresponding coherent measure of deviation is

$$\mathcal{D}(X) = \mathrm{CVaR}_\alpha(X - EX). \tag{7.9}$$

More generally, any mixture $\mathcal{R}(X) = \lambda_1 \mathrm{CVaR}_{\alpha_1}(X) + \cdots + \lambda_r \mathrm{CVaR}_{\alpha_r}(X)$ with positive weights adding to 1 gives an averse, coherent measure partnered with the deviation measure

$$\mathcal{D}(X) = \lambda_1 \mathrm{CVaR}_{\alpha_1}(X - EX) + \cdots + \lambda_r \mathrm{CVaR}_{\alpha_r}(X - EX), \tag{7.10}$$

which therefore is coherent also.

7.6. Aversity of Risk Measures of Penalty Type

The coherent risk measure $\mathcal{R}(X) = EX + \lambda \| \max\{0, X - EX\} \|_p$ for any $\lambda > 0$ and any $p \in [1, \infty]$ is averse. That is clear from the definition of strictness through (R6): We do have $\mathcal{R}(X) - EX > 0$ unless X is constant. The corresponding coherent measure of deviation is

$$\mathcal{D}(X) = \lambda \| \max\{0, X - EX\} \|_p. \tag{7.11}$$

7.7. Aversity of the Worst-Case Risk Measure

The coherent risk measure $\mathcal{R}(X) = \sup X$ is averse, again directly through the observation that it satisfies (R6): Except when X is constant, we always have $\sup X > EX$. This can also be viewed as a special case of the previous example because $\| \max\{0, X - EX\} \|_\infty = \sup X - EX$. We see then as well that the corresponding coherent measure of deviation is

$$\mathcal{D}(X) = \sup X - EX. \tag{7.12}$$

7.8. Aversity of Distributed Worst-Case Risk Measures

The coherent risk measure in (6.7) is averse, with

$$\mathcal{D}(X) = -EX + \lambda_1 \sup_{\omega \in \Omega_1} X(\omega) + \cdots + \lambda_r \sup_{\omega \in \Omega_r} X(\omega). \tag{7.13}$$

7.9. Aversity of Log-Exponential Risk Measures

The coherent risk measure in the extended sense given by $\mathcal{R}(X) = \lambda \log E[e^{X/\lambda}]$ is averse. Direct verification through (R6) works again: Since $EX = \lambda \log E[e^{[EX]/\lambda}]$, having $\mathcal{R}(X) > EX$ amounts to having $E[e^Y] > e^{EY}$ for $Y = X/\lambda$, and that is Jensen's inequality for a nonconstant Y and the strictly convex function $t \mapsto e^t$. We conclude that a coherent measure of deviation in the extended sense is furnished by

$$\mathcal{D}(X) = \lambda \log E[e^{[X - EX]/\lambda}] \quad \text{for any } \lambda > 0. \tag{7.14}$$

7.10. Another Example of a Deviation Measure in the Extended Sense

A deviation measure of a type related to so-called robust statistics is defined in terms of a parameter $s > 0$ by

$$\mathcal{D}(X) = \begin{cases} \sigma^2(X) & \text{if } \sigma(X) \le s, \\ s^2 + 2s[\sigma(X) - s] & \text{if } \sigma(X) \ge s. \end{cases}$$

Here $\sigma(X)$ could be replaced by $\mathcal{D}_0(X)$ for any deviation measure $\mathcal{D}_0$.

8. Characterizations of Optimality

For a problem of optimization in the form of (3.1) with each $\mathcal{R}_i$ a coherent measure of risk, how can solutions $\bar{x}$ be characterized? This topic could lead to a major discussion, but here we only have space for a few words. Basically this requires working with subgradients of the functions $\bar{c}_i$ in the sense of convex analysis, and that means being able to determine the subgradients of the functionals $\mathcal{R}_i$. That has been done in Rockafellar et al. [18]. The answer, for the case of $\mathcal{R}_i$ being positively homogeneous, is given in terms of the corresponding risk envelopes $\mathcal{Q}_i$. The set of subgradients Y of $\mathcal{R}_i$ at X is

$$\partial R_i(X) = \arg\max_{Q \in \mathcal{Q}_i} E[XQ]. \tag{8.1}$$

Details of what that means for various examples are provided in Rockafellar et al. [18]. Fundamentally, Lagrange multipliers, duality, and other important ideas in convex optimization revolve around the risk envelopes when invoked in the context of uncertainty.

Note, for instance, that if the random variable $\underline{c}_0(x)$ is staircased as in (2.11) and constraints of the form $\mathrm{CVaR}_{\alpha_k}(\underline{c}_0(x) - d_k) \le 0$ are imposed to tune its distribution, a Lagrangian expression in the form

$$\mathrm{CVaR}_{\alpha_0}(\underline{c}_0(x)) + y_1[\mathrm{CVaR}_{\alpha_1}(\underline{c}_0(x)) - d_1] + \cdots + y_q[\mathrm{CVaR}_{\alpha_q}(\underline{c}_0(x)) - d_q] \tag{8.2}$$

is generated in which minimization in x for fixed nonnegative multipliers $y_1, \ldots, y_1$ corresponds to minimization of $\mathcal{R}(\underline{c}_0(x))$ for the mixed CVaR risk measure

$$\mathcal{R} = \lambda_0 \mathrm{CVaR}_{\alpha_0} + \lambda_1 \mathrm{CVaR}_{\alpha_1} + \cdots + \lambda_q \mathrm{CVaR}_{\alpha_q} \tag{8.3}$$

in which the coefficient vector $(\lambda_0, \lambda_1, \ldots, \lambda_q)$ is obtained by rescaling $(1, y_1, \ldots, y_q)$ so that the coordinates add to 1. Duality, in the framework of identifying the multipliers which yield optimality, must in effect identify the weights in this mixture and therefore an implicit risk profile for the optimizer who imposed the staircase constraints.

Acknowledgments

This research was partially supported by National Science Foundation grant DMI 0457473, Percentile-Based Risk Management Approaches in Discrete Decision Making Problems.

References

[1] C. Acerbi. Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking and Finance* 26:1505–1518, 2002.

[2] C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking and Finance* 26:1487–1503, 2002.

[3] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Thinking coherently. *Risk* 10:68–91, 1997.

[4] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance* 9:203–227, 1999.

[5] A. Ben Tal and M. Teboulle. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*. 17:449–476, 2007.

[6] F. Delbaen. Coherent risk measures on general probability spaces. Working paper, Eidgenössische Technische Hochschule, Zürich, Switzerland, http://www.math.ethz.ch/~delbaen/ftp/preprints/RiskMeasuresGeneralSpaces.pdf, 2000.

[7] H. Föllmer and A. Schied. *Stochastic Finance*. Walter de Gruyter, Berlin, Germany, 2002.

[8] R. Koenker. *Quantile Regression*. Econometric Society Monograph Series, Cambridge University Press, West Nyack, NY, 2005.

[9] R. Koenker and G. W. Bassett. Regression quantiles. *Econometrica* 46:33–50, 1978.

[10] G. Pflug. Some remarks on the value-at-risk and the conditional value-at-risk. S. Uryasev, ed. *Probabilistic Constrained Optimization: Methodology and Applications*. Kluwer Academic Publishers, Norwell, MA, 2000.

[11] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.

[12] R. T. Rockafellar. *Conjugate Duality and Optimization*, No. 16 in *Conference Board of Math. Sciences Series*. SIAM, Philadelphia, PA, 1974.

[13] R. T. Rockafellar and S. P. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk* 2:21–42, 2000.

[14] R. T. Rockafellar and S. P. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance* 26:1443–1471, 2002.

[15] R. T. Rockafellar and R. J-B Wets. *Variational Analysis*, No. 317 in the series *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, Germany, 1997.

[16] R. T. Rockafellar, S. Uryasev, and M. Zabarankin. Generalized deviations in risk analysis. *Finance and Stochastics* 10:51–74, 2006.

[17] R. T. Rockafellar, S. Uryasev, and M. Zabarankin. Deviation measures in risk analysis and optimization. Research Report 2002–7, Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL, 2002.

[18] R. T. Rockafellar, S. Uryasev, and M. Zabarankin. Optimality conditions in portfolio analysis with general deviation measures. *Mathematical Programming, Series B* 108:515–540, 2006.

[19] R. T. Rockafellar, S. Uryasev, and M. Zabarankin. Master funds in portfolio analysis with general deviation measures. *Journal of Banking and Finance* 30:743–778, 2006.

[20] A. Ruszczyński and A. Shapiro. Optimization of convex risk functions. *Mathematics of Operations Research* 31:433–452, 2006.

[21] A. Ruszczyński and A. Shapiro. Conditional risk mappings. *Mathematics of Operations Research* 31:544–561, 2006.

[22] R. J-B Wets. Stochastic programming. G. Nemhauser and A. Rinnnooy Kan, eds. *Handbook of Operations Research and Management Science*, Vol. 1, *Optimization*. Elsevier Science Publishers, Amsterdam, The Netherlands, 573–629, 1987.

[23] A. A. Trindade, S. Uryasev, A. Shapiro, and G. Zrazhevsky. Financial prediction with constrained tail risk. *Journal of Banking and Finance*. Forthcoming.