ENGINEERING DECISIONS UNDER RISK-AVERSENESS*

R. Tyrrell Rockafellar
Department of Mathematics
University of Washington
rtr@uw.edu

Johannes O. Royset
Operations Research Department
Naval Postgraduate School
joroyset@nps.edu

Abstract. Engineering decisions are invariably made under substantial uncertainty about current and future system cost and response, including cost and response associated with low-probability, high-consequence events. A risk-neutral decision maker would rely on expected values when comparing designs, while a risk-averse decision maker might adopt nonlinear utility functions or failure probability criteria. The paper shows that these models for making decisions are related to a framework of risk measures that includes many possibilities. We describe how risk measures provide an expanded set of models for handling risk-averse decision makers. General recommendations for selecting risk measures lead to decision models for risk-averse decision making that comprehensively represent risks in engineering systems, avoid paradoxes, and accrue substantial benefits in subsequent risk, reliability, and cost optimization. The paper provides an overview of the framework of decision making based on risk measures.

Keywords: Risk-averse decision making, risk measures, engineering design, reliability-based optimization, utility theory, failure probability

Date: January 14, 2015

1 Introduction

Engineering centers on estimating immediate and life-cycle costs, future system responses, levels of damage, rates of deterioration, and numerous other quantities of interest. Invariably, these quantities are uncertain due to unknown environmental conditions, load patterns, as well as cost and effect of repair and retrofit actions, but their probability distributions and exceedance probability curves can often be estimated by various models; see for example Yang et al. (2009); Mahsuli and Haukaas (2013).

*This material is based upon work supported in part by the U. S. Air Force Office of Scientific Research under grants FA9550-11-1-0206 and F1ATAO1194GOO1.

Under these circumstances, engineers need to determine whether a system is adequately safe, reliable, and inexpensive as compared to given requirements. When comparing multiple designs and plans, it becomes essential to identify the safest, most reliable, and least costly design, maintenance plan, and retrofit action. In this paper, we discuss how to utilize distributional information about quantities of interest, obtained from models, to address these issues in situations of *risk averseness*.

Within a probabilistic framework, it is beneficial to view a quantity such as life-cycle cost, displacement, and damage as a random variable with distribution estimated by (stochastic) models. The difficulty arises then from the multitude of possible realizations of the random variable, which makes comparison with code, regulatory, and owner-specified requirements difficult. In fact, the development of such requirements faces the same challenge because the meaning of a requirement "at most \$100 million in damage over the lifetime of the system" is ambiguous due the random nature of the damage cost. Should the "at most" be considered over all possible realizations of the cost random variable, i.e., for all possible future scenarios of loads and structural deterioration? Alternatively, would it suffice that the damage is "adequately" or "mostly" below \$100 million, allowing for higher values in some rare scenarios? A similar situation of ambiguity arises during comparison between competing designs, retrofit plans, and other possible actions. It might be unclear if a design with low expected cost, but large variability in possible levels of cost, is preferable to a design with a somewhat higher expected cost that has little variability in its cost. Still an engineer needs to make choices and recommendations prior to learning the "true" cost or response that eventually will be realized.

The *risk-neutral* approach, focusing exclusively on the expected value of such a random variable, provides a first, natural way of addressing the issues of "adequately safe, reliable, and inexpensive" and "safest, most reliable, and least costly." For example, if a random variable Y gives the total (life-cycle) cost of a system, which is required not to exceed a threshold c, then "adequately inexpensive" would mean that the expectation $E[Y] \leq c$. Similarly, for two candidate designs of the system with associated costs, the least costly will be the design with smaller expected cost. This emphasis on expected costs is advisable for decision makers unconcerned about the possibility of high cost or excessive response. Only average cost or response is considered. Other decision makers may be risk averse and be willing to forego a low expected cost and instead select a design with, for example, low variability or low failure probability; see Stewart et al. (2011); Cha and Ellingwood (2012, 2013); Friedman et al. (2014) and references therein for recent examples and arguments for risk averseness. Of course, one can argue that some of these decision makers *should be* risk neutral. In fact, a decision maker might be risk averse for a multitude of reasons. Some reasons might be poorly justified, but others could be driven by regulatory requirements and an inability or unwillingness to model certain consequences, which might be harder to overcome. In this paper, we do not give specific recommendations about decision models, which we believe often are highly situational dependent. In particular, we avoid giving recommendations about when risk-averse models are preferred over risk-neutral models. The paper does not aim to provide a complete, or even partial, list of circumstances under which risk averseness should be adopted. We simply provide engineers with an expanded set of tools for handling risk averseness regardless how it might arise. In contrast to the risk-neutral case, risk averseness can be represented in several ways and therefore is much more challenging to model.

Of course, the issues we raise here are closely related to the numerical representations of preferences, whose modern treatment was pioneered by von Neumann and Morgenstern (1944), and go much beyond engineering to economics, finance, operations research, and other fields. In fact, we draw on these fields to introduce a set of decision models that enable engineers to assess, compare, and optimize random

variables. We make no attempt to present a comprehensive picture of this immense field; see Ditlevsen and Madsen (1996); Nathwani et al. (1997); Murphy and Gardoni (2006, 2012); Cha and Ellingwood (2012, 2013); Faber and Lind (2013); Faber et al. (2013) for engineering perspectives on risk modeling and design philosophy. Instead, we focus on *risk measures* as a main tool for building decision models. As will become abundantly clear, risk measures provide a tremendously broad class of models that relates to common choices such as simply letting random variables be represented by their expected values or by their probability of exceeding a certain threshold. Expected utility theory and dual utility theory also lead to decision models that utilize certain risk measures. Decision models based on risk measures are supported by a well-developed theory (Artzner et al., 1999; Ruszczynski and Shapiro, 2006; Ben-Tal and Teboulle, 2007; Rockafellar, 2007; Krokhmal et al., 2011; Rockafellar and Uryasev, 2013; Ruszczynski, 2013) and extensive use in financial engineering (Föllmer and Schied, 2004; Dowd, 2005) and increasingly in other fields (Commander et al., 2007; Rockafellar and Royset, 2010; Minguez et al., 2013). They are also closely connected to stochastic dominance and situations with model uncertainty. In this paper, we lay out for the first time, with a number of simple examples, the possibilities and implications of risk measures in reliability and structural engineering. The framework of risk measures provides an expanded set of decision models for handling risk-averse decision makers beyond the usual choices of failure probabilities, expected utility theory, and its various extensions. Among the insights, we show how distributional uncertainty leads to risk averseness of a specific kind.

One might naturally question why failure probabilities and expected utility theory do not suffice as decision models for risk averse situations. We provide arguments why the failure probability, despite its popularity, suffers from several shortcomings, with a specific risk measure, the superquantile risk measure, furnishing a viable alternative. We show that "regular" measures of risk in some sense probe deeper than expected utility theory by considering how one can mitigate poor performance. Moreover, risk measures are supported by dual utility theory (Yaari, 1987), which relies on a set of axioms that are weaker than those of expected utility theory. Consequently, risk measures appear as viable alternatives to expected utility theory.

A choice of risk measure, and more generally decision model, must necessarily be driven by the need of a specific application and the (financial) resources, responsibilities, and attitudes of the decision maker. We therefore shy away from universally recommending one risk measure. Instead we address the situation when a risk-averse model is needed and lay out the vast possibilities found within the framework of risk measures, describe connections with utility theory, provide criteria "reasonable" risk measures should satisfy, and highlight the difficulties and paradoxes that may arise when these criteria are violated.

We stress that "good" decisions not only rely on the selection of an appropriate decision model, but also on appropriate modeling of the cost, damage, response, and other quantities of interest. For example, when the focus is on cost, it is usually important to develop appropriate probabilistic models of the cost of consequences that go beyond the immediate cost and capture indirect cost. In fact, Section 3.5 gives a specific explanation of how risk averseness arise from incomplete and inaccurate probabilistic models. In a highly precise manner, we show the value of improved consequence modeling in many contexts. We stress that risk-averse decision models should not be used as a substitute for proper consequence modeling. Engineers should aim to go as far as possible in developing accurate probabilistic models of all relevant factors. It is in view of such a set of probabilistic models that engineers, decision makers, and other involved parties should select a decision model.

Although our perspective of modeling future cost, level of response, and other quantities as a random

variable captures many practical situations, it ignores the often intricate and gradual revelation of the "true" cost and response over time, usually intertwined with decisions about maintenance and retrofit. Modelling of this situation fully would lead to multi-stage risk-averse decision making and the need for additional mathematical machinery such as those related to stochastic processes; see for example Ruszczynski (2013). To avoid obscuring the central tenets of the paper, we therefore focus exclusively on the situation where we "today" must make an assessment and decision about a system with an uncertain (future) cost or response, modelled as a random variable and typically associated with some time period. The random variable will be realized only *after* the decision is made. In this case, there will be no opportunities for optimizing corrective actions after some of the uncertainty has been resolved. Though, as we see in Section 3, "regular" risk measures implicitly capture a two-stage decision process. The general case leads to similar developments, but is avoided here for conciseness of exposition.

We proceed in Section 2 with a description of the situation where a decision must be made in the presence of uncertain cost and response. Section 3 introduces measure of risk and shows that most decision models can be interpreted in terms of such measures. We also provide simple examples as well as supporting theory. These foundations allow us to address the issue of how to optimally make decisions under uncertainty in Section 4. The paper ends with conclusions in Section 5. For completeness, we provide background material in an appendix.

2 Decisions under Risk-Averseness

We consider the situation where one or more (stochastic) models characterize the (future) "performance" of a system in terms of a random variable Y that describes *response quantities* such as displacement, stress, capacity exceedance, and crack size, or *costs* such as construction cost, maintenance cost, damage cost, repair cost, and life-cycle cost. For example, a bar with random cross-section A, random yield stress $\sigma^{\max}$, and random load S may have $Y = S - \sigma^{\max}A$, which describes the difference between load and capacity. Alternatively, Y could be the total life-cycle cost of a bridge or offshore structure, including direct and indirect costs of "failures," described by a series of complex models. Although the framework discussed below does not require Y to be a "good" model of reality, it is practically important that Y incorporates all important aspect of the underlying phenomena such as, for example, second and third degree consequences of failures. We adopt the convention that high realizations of Y are undesirable, which is natural as Y often represents monetary cost. We maintain the convention in the case of response quantities also, which at most requires minor adjustments. For instance, in the simple example above one needs to consider $S - \sigma^{\max}A$ instead of $\sigma^{\max}A - S$. The exact meaning of Y varies of course with the application. In practice, multiple response quantities and types of costs could be under consideration simultaneously. Section 4 provides formulations of problems dealing with that situation, but the majority of the paper discusses a single response or cost for simplicity of exposition. Regardless of the situation, a response or cost random variable Y and its (estimated) distribution serve as the basis for decisions regarding the system. For engineers, regulators, code-developers, and system owners, the following questions then arise:

$$\text{How "risky" is } Y? \qquad \text{Will } Y \text{ be "adequately" } \leq c?$$

Here the constant c may be a budgeted amount of money or a threshold that a response must stay below to avoid "failure." When comparing a candidate design that results in the response or cost random

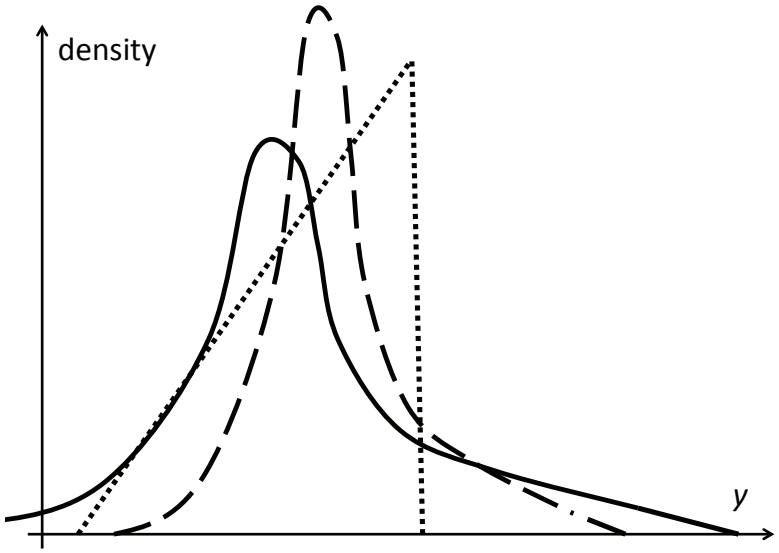

Figure 1: Probability densities for various cost random variables

variable Y with a benchmark or alternative design represented by random variable Y', we may ask:

Is the design with Y preferable to that with Y'?

These fundamental questions are best illustrated by an example.

Examples of cost densities. We consider three alternative designs with costs described by the probability densities in Figure 1. (An interpretation in terms of system responses would follow similarly.) The density given by a solid line has the lowest expected value and, under an expected-value criterion, would be the preferred design. However, the dashed design has smaller variance at the expense of a larger expected value. Depending on a decision maker's preference, the benefits of a reduced variance may more than outweigh the increased expected value. The dotted density has the same expected value and variance as the dashed one, but the possibility of high costs is much smaller. A decision maker may want to go beyond variance and look at variability in an asymmetric manner with deviations from the mean on the high side viewed more unfortunate than deviations on the low side. With this perspective, the dotted density may emerge as the "best."

Decision models provide means for addressing the above fundamental questions in a systematic way. Although the questions may at first appear distinct, with "adequately below c" leading to constraints and "preferences" giving rise to objective functions in design optimization problems, the difference is insignificant. The task of finding a design with cost below c is essentially equivalent to minimizing the cost and then checking whether the minimum value is below c. The task of minimizing cost is also achieved by finding a design with cost below a threshold c and then repeating the process for successively smaller c. Consequently, there is no valid reason for treating one question differently than the other. Both cases require a quantification of the "risk" in a random variable Y.

The simplest decision model is to rely on expected values exclusively. As eluded to in Section 1, the questions are then simply answered by replacing the random variables by their expectations and then comparing the resulting numbers. The justification for such an approach is well-known: A decision maker that will see many events, each causing an independent realization of a random variable, will face an *average* cost or response per event that is close to the expectation of the random variable. A design with the smallest average cost or response would then be "best," on average. A governmental-type decision maker dealing with monetary cost is often believed to fall in this category due to its large portfolio of structures and/or its long time horizon. We believe, however, that there are situations where risk-averse approaches might be considered; see Stewart et al. (2011); Cha and Ellingwood (2012, 2013) for such arguments. In particular, if the focus is on a response quantity Y such as stress level, displacement, or crack size, then it would make little sense to impose the condition that $E[Y] \leq c$ for some threshold c. To have tolerable response only on average is usually unacceptable and, of course, a risk-averse approach using failure probability is in many areas standard; see for example Ditlevsen and Madsen (1996). However, as we see in the next section, this is only one of several possible quantities on which to base a risk-averse decision model. Also, as we see in Section 3.5,

> a risk-neutral decision maker uncertain about the distribution of costs is equivalent to a risk-averse decision maker certain about the distribution of costs.

Consequently, an otherwise risk-neutral decision maker may be driven to risk averseness due to perceived incompleteness of cost models. Risk measures help to explain this phenomenon.

Decision models provide the linkage between descriptive models of system response and cost, and actual decision making. They deal with comparison of response and cost random variables. We therefore need to consider a family of random variables corresponding to different design or alternatives. We let $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ be a vector describing such alternatives, where the components may specify member sizes, layout, material choices, maintenance schedules, retrofit actions, etc. The random variable $Y(\boldsymbol{x})$ then specifies the (random) response or cost, given design $\boldsymbol{x}$. If the response or cost is given in terms of a function g that depends on the design $\boldsymbol{x}$ and a random vector $\boldsymbol{V}$ describing uncertain parameters, then $Y(\boldsymbol{x}) = g(\boldsymbol{x}, \boldsymbol{V})$. For example, a bar with random load S and yield stress $\sigma^{\max}$, and cross section A, to be selected by the engineer, might have $\boldsymbol{V} = (S, \sigma^{\max})$, $\boldsymbol{x} = A$, and $Y(\boldsymbol{x}) = g(\boldsymbol{x}, \boldsymbol{V}) = S - \sigma^{\max} A$. Then, the random variable $Y(\boldsymbol{x})$ describes the random load exceedance for design $\boldsymbol{x} = A$; positive realizations imply that the load exceeds the capacity and nonpositive realizations lead to satisfactory performance. When the discussion centers on a single random variable, we drop $(\boldsymbol{x})$ and simply write Y as before.

3 Risk Measures

A broad class of decision models that encapsulates essentially all reasonable models rely on measures of risk as defined next:

> A *measure of risk* is a functional $\mathcal{R}$ that assigns to a response or cost random variable Y a number $\mathcal{R}(Y)$, which could be infinity, as a quantification of the risk in Y.

The answer to the question of how "risky" is Y, is now simply defined to be $\mathcal{R}(Y)$. We say that Y is "adequately" $\leq c$ if $\mathcal{R}(Y) \leq c$. The comparison of two choices Y and Y' then reduces to comparing

$\mathcal{R}(Y)$ and $\mathcal{R}(Y')$; see Krokhmal et al. (2011); Rockafellar and Uryasev (2013) for recent reviews of risk measures. We start with examples of risk measures that illustrate the breadth of the framework as well as its connections to classical decision models. We then proceed with axioms that "good" risk measures should satisfy. The section ends with discussion of deeper connections with dual utility theory, expected utility theory, and cases with model uncertainty.

3.1 Examples

Since a risk measure simply converts a response or cost random vector into a scalar, a large number of possibilities exist. We next describe some choices; see Rockafellar and Uryasev (2013) for additional examples.

Expectation. The choice $\mathcal{R}(Y) = E[Y]$, the expected value, is simple, but not sensitive to the possibility of high values; see the discussion of Figure 1. Obviously, this choice incorporates no level of risk averseness.

Worst-case. The choice

$$\mathcal{R}(Y) = \sup Y = \text{ the smallest value that } Y \text{ exceeds only with probability zero}$$

is conservative, usually overly so as it is infinite for distributions such as the normal and exponential. In fact, the corresponding decision model ignores all the information in the distribution of Y except its highest "possible" realization. Still, in some applications there may be thresholds that simply should not be exceeded as discussed in Murphy and Gardoni (2012).

Failure probability. The failure probability

$$p(Y) = \text{prob}(Y > 0),$$

assuming that positive values are considered "failure," is widely used in reliability analysis; see for example Ditlevsen and Madsen (1996, Chapter 12) for a discussion. Recent proposals centered on the capabilities of individuals after disasters also rely on probabilities of exceeding thresholds specifying essential needs (Murphy and Gardoni, 2006, 2008; Gardoni and Murphy, 2010; Murphy and Gardoni, 2012). However, there are two immediate concerns with failure probability models. (A third surfaces below that in fact makes us shy away from calling the failure probability a "regular" measure of risk.) First, there may be two design $\boldsymbol{x}$ and $\boldsymbol{x}'$ with the same failure probability, i.e., $p(Y(\boldsymbol{x})) = p(Y(\boldsymbol{x}'))$, but the distributions of $Y(\boldsymbol{x})$ and $Y(\boldsymbol{x}')$ may be different, especially in the critical upper tail. For example, if $Y(x) = g(x, \boldsymbol{V}) = 100 - xV_1 - (1-x)V_2$, where V_1 is normally distributed with mean 150 and standard deviation 15 and V_2 is triangularly distributed on the range $[98.40, 175.8]$, with mode at 98.40, then the two designs $x = 1$ and $x = 0$ result in rather different distributions $Y(0) = g(0, \boldsymbol{V})$ and $Y(1) = g(1, \boldsymbol{V})$. Figure 2 shows the tails of the corresponding densities. It is clear that design $x = 1$ allows much larger values of $Y(x)$, which may be of concern for a decision maker as it indicates the possibility for more dramatic failures. Still, one can show that the designs have identical failure probability. Consequently, the failure probability is insensitive to the tail of the distribution and an exclusive focus on the corresponding decision models may hide significant risks. The second concern when using

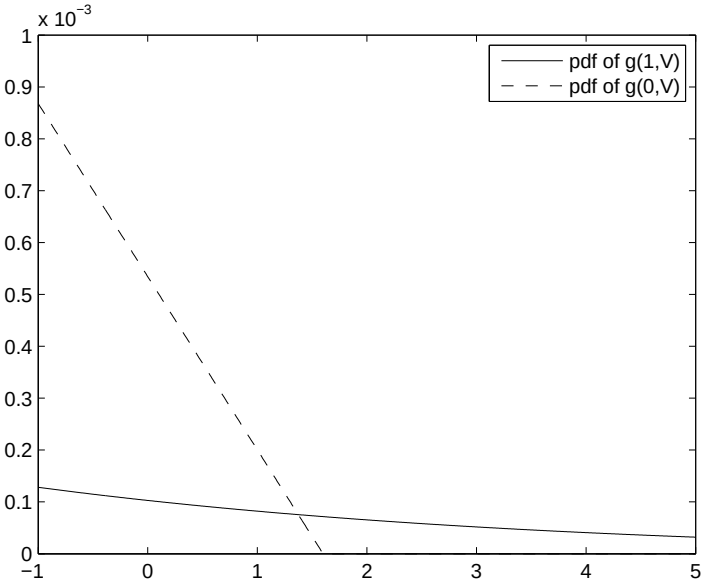

Figure 2: Tails of the densities of $Y(1) = g(1, \boldsymbol{V})$ and $Y(0) = g(0, \boldsymbol{V})$

the failure probability is its lack of convexity and smoothness as a function of the design vector $\boldsymbol{x}$. This deficiency dramatically increases the difficulty of solving design optimization problems involving $p(Y(\boldsymbol{x}))$. The absence of convexity makes it fundamentally harder to obtain globally optimal designs as the standard, efficient algorithms can only be guaranteed to return locally optimal designs. The lack of smoothness makes gradients of the failure probability with respect to the design $\boldsymbol{x}$ unavailable and one is forced to rely on slower derivative-free algorithms. A solution of such optimization problems also becomes unstable under perturbations and therefore further draws into doubt the viability of the model; see Section 4 and also Rockafellar and Royset (2010) for further details. Superquantile risk measures described below are viable alternatives to failure probability models that avoid these difficulties.

Quantile. For $\alpha \in (0, 1)$, the α-quantile of a random variable Y, $q_\alpha(Y)$, is simply $F_Y^{-1}(\alpha)$ when the cumulative distribution function F_Y of Y is strictly increasing; see Figure 3. More generally,

$$q_\alpha(Y) = \text{ the smallest } y \text{ that has } F_Y(y) \text{ no smaller than } \alpha.$$

The choice of risk measure

$$\mathcal{R}(Y) = q_\alpha(Y)$$

is widely used in financial engineering under the name "value-at-risk" with typically an α of nearly one, and is equivalent to the failure probability measure of risk. Specifically,

$$p(Y) \leq 1 - \alpha \text{ if and only if } q_\alpha(Y) \leq 0. \tag{1}$$

Here we note that if Y is greater than zero with probability one, then certainly $q_\alpha(Y) > 0$ for any $\alpha \in (0, 1)$ and the failure probability is one. Consequently, the relationship still holds in this extreme case.

Utility. For a utility function u, the classical expected utility theory of von Neumann and Morgenstern (1944) considers $E[u(Y)]$. We avoid calling this a risk measure as our orientation is different than that of expected utility theory, which prefers higher values over lower ones. Our setting focuses on cost and response, where lower values are desirable. Although this might be overcome by inserting various minus signs, a more serious issue is that expressions of this kind will generally not be "regular" as explained in Section 3.2. However, we make connections with expected utility and risk measures in Section 3.4. Of course, if $u(y) = ay + b$ for some scalars $a > 0$ and b, then preferences represented by u is equally well represented by the expected value. Specifically, $E[Y] < E[Y']$ if and only if $E[u(Y)] < E[u(Y')]$. Consequently, under a linear or affine utility function, decisions based on expected utility are essentially equivalent to those based on expected values.

Safety margin. A natural choice motivated by statistical confidence intervals is to set

$$\mathcal{R}(Y) = E[Y] + \lambda SD[Y],$$

where $SD[Y]$ is the standard deviation of Y and λ a positive constant. Here the risk includes a notion of variability, but does so in a symmetrical manner. Large variability on the high side can remain undetected if it is compensated by small variability on the low side. For example, in Figure 1 the dotted and dashed densities have the same expected value and standard deviation. Consequently, this risk measure would assign the two cases the same risk. Figure 1 shows, however, that the possibility of high costs is much larger for the dashed density, which could be disconcerting.

Superquantile. The α-superquantile of Y at probability $\alpha \in (0, 1)$ is given by

$$\bar{q}_\alpha(Y) = \frac{1}{1-\alpha} \int_\alpha^1 q_\beta(Y) d\beta, \tag{2}$$

i.e., an α-superquantile is an average of quantiles for probability levels $\alpha < \beta < 1$ (Acerbi, 2002). When the cumulative distribution function of Y has no discontinuity at the realization $y = q_\alpha(Y)$, we have the equivalent formula

$$\bar{q}_\alpha(Y) = E[Y \mid Y \geq q_\alpha(Y)]$$

i.e., the α-superquantile is simply the conditional expectation of Y above the α-quantile as illustrated in Figure 3. Despite its somewhat complicated definition, convenient expressions facilitate the computation of superquantiles making them almost as accessible as an expectation. If Y is normally distributed with mean μ and standard deviation σ, then

$$\bar{q}_\alpha(Y) = \mu + \frac{\sigma\varphi(\Phi^{-1}(\alpha))}{1-\alpha},$$

where φ and Φ are the probability density and cumulative distribution functions for a standard normal random variable. If Y follows a discrete distribution with realizations $y_1 < y_2 < ... < y_n$ and corresponding probabilities $p_1, p_2, ..., p_n$, then

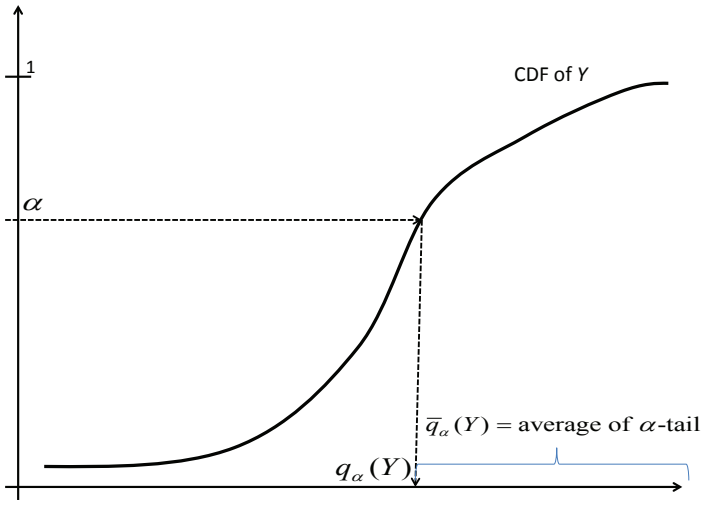

Figure 3: Illustration of quantile and superquantile of random variable Y using cumulative distribution function (CDF).

$$\bar{q}_\alpha(Y) = \begin{cases} \sum_{j=1}^{n} p_j y_j & \text{for } \alpha = 0 \\ \frac{1}{1-\alpha}\left[\left(\left(\sum_{j=1}^{i} p_j\right) - \alpha\right) y_i + \sum_{j=i+1}^{n} p_j y_j\right] & \text{for } \sum_{j=1}^{i-1} p_j < \alpha \leq \sum_{j=1}^{i} p_j < 1 \\ y_n & \text{for } \alpha > 1 - p_n. \end{cases} \tag{3}$$

We note that the realizations are sorted, without loss of generality, to simplify the formula. Generally,

$$\bar{q}_\alpha(Y) = \text{ the minimum value of } c + \frac{1}{1-\alpha} E[\max\{0, Y - c\}] \text{ across all scalars } c, \tag{4}$$

i.e., a superquantile is the minimum value of a one-dimensional convex optimization problem involving an expectation easily solved when the distribution of Y is known.

A risk measure that focuses primarily on the important upper tail of the distribution of Y is then the superquantile risk measure

$$\mathcal{R}(Y) = \bar{q}_\alpha(Y);$$

also called conditional value-at-risk (Rockafellar and Uryasev, 2000). A superquantile risk measure depends on the parameter α that represents the degree of risk averseness of the decision maker. For $\alpha = 0$, $\bar{q}_\alpha(Y) = E[Y]$ and therefore corresponds to the risk-neutral situation. An $\alpha = 1$ gives $\bar{q}_\alpha(Y) = \sup Y$ and therefore corresponds to the ultimate risk-averse decision maker. Using a collection of different values of α and let the risk measure be a linear combination of the corresponding superquantiles is also a possibility, which we return to in Section 3.3.

The correspondence between a failure probability constraint $p(Y) \leq 1-\alpha$ and the quantile condition $q_\alpha(Y) \leq 0$ is elaborated above. Analogously, a superquantile condition $\bar{q}_\alpha(Y) \leq 0$ corresponds to the

condition $\bar{p}(Y) \leq 1-\alpha$, where $\bar{p}(Y)$ is the buffered failure probability of Y defined as the probability α that satisfies $\bar{q}_\alpha(Y) = 0$. We refer to Rockafellar and Royset (2010) for a discussion of the advantages that emerge from replacing a failure probability by a buffered failure probability. A viable and in fact conservative alternative to a failure probability condition of the form $p(Y) \leq 1-\alpha$ is therefore a buffered failure probability requirement $\bar{p}(Y) \leq 1-\alpha$ or equivalently the superquantile risk requirement $\bar{q}_\alpha(Y) \leq 0$. The prominence of superquantiles becomes apparent in the next subsection where we discussion properties required for "good" measures of risk.

3.2 Coherency and Regularity

With the abundance of possible risk measures (and we refer to Rockafellar and Uryasev (2013) and references therein for many more), there is a demand for guidance on what would constitute a good and useful measure of risk. There are two concepts that stand out in this regards: coherency and regularity. We discuss each in turn.

A measure of risk $\mathcal{R}$ is *coherent* in the sense of Artzner et al. (1999)[†] (see also Delbaen (2002)) if it satisfies the following axioms:

$$
\begin{aligned}
&\mathcal{R}(Y) = c \text{ for constant random variables } Y \equiv c \text{ (constant equivalence)},\\
&\mathcal{R}(Y) \leq \mathcal{R}(Y') \text{ when } Y \leq Y' \text{ with probability one (monotonicity)},\\
&\mathcal{R}(Y+Y') \leq \mathcal{R}(Y) + \mathcal{R}(Y') \text{ (subadditivity)},\\
&\mathcal{R}(\lambda Y) = \lambda \mathcal{R}(Y) \text{ for } \lambda > 0 \text{ (positive homogeneity)}.
\end{aligned}
$$

The constant equivalence simply requires that the risk of a random variable that is actually a deterministic constant is the value of that constant. The monotonicity says that we would deem Y' no less risky than Y if every realization of Y' is no smaller than the corresponding realization of Y, with the possible exception for an event with probability zero. Subadditivity expresses the requirement that combining two random variables should not increase risk. The left-hand side in the expression gives the risk of the combined cost or response, which of course depends on the joint distribution of Y and Y'. The right-hand side gives the sum of the risks for the two random variables separately, only relying on their marginal distributions. Positive homogeneity expresses the desire to have invariance under scaling. If the units of Y is converted from one currency to another or from Pa to psi, then the risk is also simply scaled correspondingly.

Coherency holds for the choices $\mathcal{R}(Y) = E[Y]$, $\mathcal{R}(Y) = \sup Y$, and $\mathcal{R}(X) = \bar{q}_\alpha(Y)$, but it is absent in general for $E[u(Y)]$ and $p(Y)$, because constant equivalence and positive homogeneity fail. In fact, a utility function distorts a deterministic Y in a possibly undesirable manner: If Y is a deterministic constant, then a coherent $\mathcal{R}(Y)$ equals that constant always. But, $E[u(Y)]$ could be far from that constant in general. Moreover, $E[u(Y)]$ is sensitive to thresholds in the definition of Y (for example through a limit-state function). Specifically, $E[u(Y+c)] \neq E[u(Y)] + c$ in general. In contrast, a coherent risk measure changes linearly with threshold, i.e., $\mathcal{R}(Y+c) = \mathcal{R}(Y) + c$ for any constant c. Similarly, the lack of positive homogeneity in utility-based expressions induces dependence on scaling, which is typically problematic. The situation is similar for the failure probability.

[†]The statement here is slightly different, but equivalent, to that in Artzner et al. (1999)

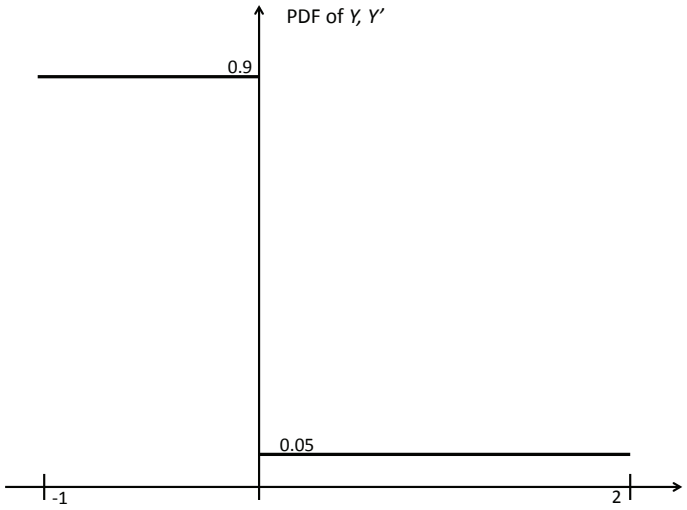

Figure 4: Density of Y and Y' in subadditivity counterexample.

Neither the choice $\mathcal{R}(Y) = E[Y] + \lambda SD(Y)$, because of lack of monotonicity, nor the choice $\mathcal{R}(Y) = q_\alpha(Y)$, because of the lack of subadditivity, is coherent. We give two examples that illustrate these shortcomings.

First, we consider the lack of monotonicity of the risk measure $\mathcal{R}(Y) = E[Y] + \lambda SD(Y)$. For two random variables Y and Y', with $\text{prob}(Y = 0, Y' = 0) = 1/2$ and $\text{prob}(Y = -2, Y' = -1) = 1/2$, clearly, $Y \leq Y'$ with probability one, $\mathcal{R}(Y) = E[Y] + 2SD(Y) = -1 + 2 \cdot 1 = 1$ and $\mathcal{R}(Y') = E[Y'] + 2SD(Y') = -1/2 + 2 \cdot 1/2 = 1/2$. However,

$$E[Y] + 2SD[Y]1 > E[Y'] + 2SD[Y'] = 1/2$$

and monotonicity fails for the risk measure $\mathcal{R}(Y) = E[Y] + \lambda SD(Y)$ with $\lambda = 2$.

Second, we consider subadditivity of the quantile risk measure. Suppose that Y and Y' quantifies the damage of bridge 1 and 2 under a future hazard, respectively. Let $\mathcal{R}(Y)$ be the amount of money deemed necessary to put aside to cover damage of bridge 1 and $\mathcal{R}(Y')$ the amount to cover damage of bridge 2. $\mathcal{R}(Y + Y')$ is the amount of money deemed necessary to put aside to cover damage at both. The following example is taken from Artzner et al. (1999). Suppose that Y and Y' are independent and identically distributed with probability density function illustrated in Figure 4. It is relatively straightforward to compute that $q_{0.9}(Y) = q_{0.9}(Y') = 0$, but $q_{0.9}(Y + Y') \approx 0.5$ and subadditive fails. In view of this discussion, we see that "natural" choices of risk measures might lead to paradoxes better avoided.

Another concept is that of regularity. A measure of risk $\mathcal{R}$ is *regular* in the sense of Rockafellar and Uryasev (2013) if it satisfies the constant equivalence axiom as well as:

$\mathcal{R}((1-\tau)Y + \tau Y') \leq (1-\tau)\mathcal{R}(Y) + \tau\mathcal{R}(Y')$ for all Y, Y' and $\tau \in (0,1)$ (convexity);
$\{Y \mid \mathcal{R}(Y) \leq c\}$ is a closed set for every constant c (closedness);
$\mathcal{R}(Y) > E[Y]$ for nonconstant Y (averseness).

The convexity axiom is closely related to subadditivity and in conjunction with positive homogeneity is in fact equivalent to that property. The closedness axiom is included for technical reasons and requires further limitation of the random variables under consideration, for example to those with finite second moments where closedness then can be defined in terms of convergence of mean squares. Without going into technical details (Rockafellar and Uryasev, 2013), we simply note that $\mathcal{R}(Y) = \sup Y$, $\mathcal{R}(Y) = E[Y] + \lambda SD(Y)$, and $\mathcal{R}(X) = \bar{q}_\alpha(Y)$, $\alpha \in (0,1)$, are regular. The choice $\mathcal{R}(Y) = E[Y]$ fails the averseness axiom and $\mathcal{R}(Y) = q_\alpha(Y)$ the convexity axiom. The expressions $E[u(Y)]$ and $p(Y)$ fail the constant equivalence axiom. The lack of coherency and regularity of the quantile (failure probability) risk measure is the third concern eluded to above when discussing the failure probability.

The coherency and regularity axioms overlap, but are not equivalent as the above examples illustrate. Regularity insists on averseness, but coherency stresses positive homogeneity. Both notions impose conditions that are natural in almost all situations and that are significantly weaker than those of expected utility theory (see the appendix). Of course, within the class of coherent and regular risk measures there is much flexibility and the axioms can simply be viewed as minimum requirements. However, risk measures based on a single or a linear combination of superquantiles are both coherent and regular, and emerge as attractive alternatives. The next subsections and Section 4 provide further context and highlight benefits derived from using coherent and regular measures of risk.

3.3 Decision Theoretic Foundations for Superquantile Risk

It is well known that the preferences of a decision maker that obeys a set of axioms is fully captured by an expected utility function; see the appendix for a review. A parallel situation exists for the large class of risk measures based on superquantiles, but now the connection is with *dual utility theory* (Quiggin, 1982; Yaari, 1987).

Dual utility theory provides a range of decision models where preferences are based on

$$D(Y) = \int_0^1 q_\beta(Y) dw(\beta) = \text{ a weighted sum of quantiles at different probability levels } \beta, \qquad (5)$$

for some bounded, nondecreasing, and continuous weighting function w called the *rank-dependent utility function*. If the rank-dependent utility function $w(\beta) = \beta$, then $dw(\beta) = d\beta$ and $D(Y) = E[Y]$, and the preferences rely on expected values only. If piecewise linear with $w(\beta) = \max\{0, \beta/(1-\alpha) - \alpha/(1-\alpha)\}$ for fixed $\alpha \in (0,1)$, then

$$D(Y) = \frac{1}{1-\alpha} \int_\alpha^1 q_\beta(Y) d\beta = \bar{q}_\alpha(Y).$$

This choice averages the quantiles in the upper tail above a probability level α and therefore coincides with the α-superquantile of Y; see (2). Other weighting schemes, i.e., other rank-dependent utility

functions, provide other risk measures $\mathcal{R}(Y) = D(Y)$, mostly consisting of coherent and regular risk measures formed by combinations of superquantiles. Consequently,

> decisions based on risk measures formed by linear combinations of superquantiles are justified by dual utility theory.

The reliance on dual utility theory is attractive as its axioms are substantially weaker than those of the classical expected utility theory as summarized in the appendix.

3.4 Regular Risk Measures from Expected Utility Functions

In the previous section, we established connections between a class of risk measures and dual utility theory. Now, we relate to classical expected utility theory. We show that regular measures of risk can be viewed as the outcome of a two-stage decision process involving a utility function providing fundamental connections also between expected utility theory and risk measures. To avoid the awkward inconsistency between our orientation concerned with high values of Y with that of utility theory, seeking high values, we define an analogous concept to a utility function in our context where lower values are preferred to higher ones.

A *measure of regret* is a functional $\mathcal{V}$ that assigns to a random variable Y a number $\mathcal{V}(Y)$, which may be infinity, as a quantification of the displeasure with the mix of possible realizations of Y. It could correspond to a utility function u through

$$\mathcal{V}(Y) = -E[u(-Y)], \tag{6}$$

but we ensure that it is anchored at zero. Hence, we insist that $\mathcal{V}(0) = 0$. In (6), this is achieved by having $u(0) = 0$.

Analogously to the regularity of risk measures, we say that a measure of regret is regular if it satisfies the closedness and convex axioms, and

$$\mathcal{V}(0) = 0, \quad \text{but } \mathcal{V}(Y) > E[Y] \text{ when } Y \text{ is not the constant zero.}$$

Consequently, every concave utility function with $u(0) = 0$ and $u(y) < y$ for $y \neq 0$ gives a regular measure of regret through (6). Examples of regular measures of regret include $\mathcal{V}(Y) = E[Y] + \lambda E[Y^2]$, with $\lambda > 0$, and $\mathcal{V}(Y) = \frac{1}{1-\alpha} E[\max\{0, Y\}]$, with $\alpha \in (0, 1)$. The latter expression is illustrated in Figure 5, where negative realizations of Y are assigned zero regret, but positive realizations are viewed increasingly "regretable," with the increase being linear. This expression corresponds to a piecewise linear utility function with a kink at zero. A larger α implies a higher degree of risk averseness.

Major advantages derive from the following fact (Rockafellar and Uryasev, 2013): A regular measure of risk $\mathcal{R}$ can be constructed from a regular measure of regret $\mathcal{V}$ through the one-dimensional optimization problem

$$\mathcal{R}(Y) = \text{ the minimum value of } c + \mathcal{V}(Y - c) \text{ across all scalars } c. \tag{7}$$

For example, a superquantile measure of risk $\mathcal{R}(Y) = \bar{q}_\alpha(Y)$ derives from the measure of regret $\mathcal{V}(Y) = \frac{1}{1-\alpha} E[\max\{0, Y\}]$, which leads to the already claimed expression (4). A large number of other measures of risk can be constructed in a similar manner; see Rockafellar and Uryasev (2013). With the connections between regret and relative utility, this implies that

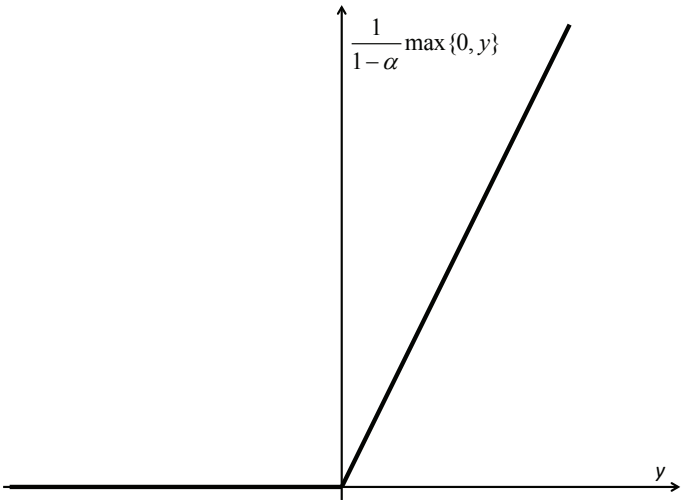

Figure 5: Illustration of regret.

every utility function u, with $u(0) = 0$ and $u(y) < y$ when $y \neq 0$, is in correspondence with a measure of risk through (6) and (7).

These connections also provide a path to determine the risk averseness for a specific decision maker. If a basic reference lottery ticket question can elicit a relative utility function for the decision maker, then the conversion of that function into regret yields a risk measure for the decision maker through (7).

The trade-off formula (7) provides important interpretations of a regular measure of risk as the result of a two-stage decision process involving a regular measure of regret (and therefore also a relative utility function). We first consider an interpretation in terms of costs. Suppose that Y gives the damage cost of a system and the measure of regret $\mathcal{V}(Y)$ quantifies our displeasure with the possible damage costs. In (7), view c as the money put aside today to cover future damage costs and $Y - c$ as the net damage cost in the future. Then, $c + \mathcal{V}(Y - c)$ becomes the total cost consisting of the sum of the money put aside today plus the current displeasure with future damage costs. The risk $\mathcal{R}(Y)$ is then simply the smallest possible total cost one can obtain by selecting the amount to put aside today in the best possible manner.

A second interpretation in terms of strength of a structure is also possible. Now suppose that Y is given by a load minus a capacity and therefore describes the possible capacity shortfall. A measure of regret $\mathcal{V}(Y)$ quantifies our displeasure with possible shortfalls. Let c in (7) be interpreted as the additional capacity added to the structure now. Then $Y - c$ is the capacity shortfall of the improved structure. Consequently, $c + \mathcal{V}(Y - c)$ sums the additional capacity c and the current displeasure with future shortfalls, and can therefore be interpreted as the current displeasure with future *effective shortfall*. A risk measure $\mathcal{R}(Y)$ is then simply the smallest possible effective shortfall obtained by optimally selecting the additional capacity implemented. Therefore, in contrast to utility theory, the use of a regular risk measure incorporates in its quantification of risk the ability to mitigate poor performance through strengthening in some sense.

With the close connection between regret (and therefore also relative utility) and risk, one may be led to believe that a decision model based on regret (or equivalently relative utility) would be equivalent to one based on the corresponding risk measure. A simple example shows that this conclusion is incorrect.

Example: Regret and risk based decisions. We consider the regret $\mathcal{V}(Y) = \frac{1}{1-\alpha}E[\max\{0, Y\}]$ and the corresponding superquantile risk measure $\mathcal{R}$ given by (7). Suppose that Y is uniformly distributed on $[a, b]$ with $a < 0 < b$. Then it is easy to show that

$$\mathcal{V}(Y) = \frac{b^2}{2(1-\alpha)(b-a)} \quad \text{and} \quad \mathcal{R}(Y) = \frac{a+b+\alpha(b-a)}{2}.$$

Now consider two possible designs with Y and Y' uniformly distributed on $[-3/2, 1]$ and $[-8, 2]$, respectively. Using the above expression and with $\alpha = 4/5$, we find that $\mathcal{V}(Y) = \mathcal{V}(Y') = 1$. Consequently, a regret-based (or equivalently a utility-based) decision model of this kind would consider the two designs equivalent. However, $\mathcal{R}(Y) = 3/4$ and $\mathcal{R}(Y') = 1$, and therefore Y is strictly better than Y' in a decision model relying on the corresponding superquantile risk measure. In some sense, the consideration of $\mathcal{R}$ instead of $\mathcal{V}$ directly provides a deeper representation of preferences as it allows for the (optimal) shifting of a threshold through (7). In this specific example, Y allows for a more effective mitigation of poor performance through a shift in threshold than Y'.

3.5 Risk-Neutrality under Distributional Uncertainty

Since regular measures of risk and regret are proper convex functionals on a space of random variables, duality through the Legendre-Fenchel transform provides important insight and alternative avenues to utilize; see for example Ruszczynski and Shapiro (2006); Pflug (2007); Bot et al. (2010); Rockafellar and Uryasev (2013). Specifically, every regular measure of risk that is positively homogeneous can be expressed in the form

$$\mathcal{R}(Y) = \text{ the maximum value of } E[YQ] \text{ across } Q \in \mathcal{Q}, \tag{8}$$

where Q is a random variable that is taken from a set $\mathcal{Q}$ of random variables called a risk envelope associated with the risk measure. For example, if $\mathcal{R}(Y) = \bar{q}_\alpha(Y)$, then $\mathcal{Q}$ consists of those random variables with realizations between zero and $1/(1-\alpha)$ and that has expectation one; see Rockafellar and Uryasev (2013); Ruszczynski (2013) and references therein for background.

Since in the positively homogeneous and monotonic case every risk envelope must exclusively contain nonnegative random variables with expectation one, the expression can be interpreted as a change of probability distribution for Y. Specifically, $\mathcal{R}(Y)$ becomes the worst-case expectation of the random variable over a set of possible changes in distributions. We see therefore that risk measures help explain how risk averseness might arise. For example, in the case of a superquantile risk measure and a discrete probability distribution of Y, we obtain the interpretation that it is simply the worst-case expectation of Y when considering all probability distribution changes such that the probability of a realization is scaled with a number between zero and $1/(1-\alpha)$. Consequently, the use of a regular measure of risk builds in robustness to uncertainty in the distribution of Y for a decision maker centered on an expectation decision model. A specific example helps illustrate this further.

Example: Uncertainty about distribution. We consider the simple situation where the random variable Y of a system takes the value 1 with probability 0.1 and the value 0 with probability 0.9, with expected value 0.1. A risk-neutral decision maker centered on the expectation would use 0.1 in numerical comparisons with other systems and requirements. Next, we consider a risk-averse decision maker that has adopted the superquantile risk measure with $\alpha = 0.8$. Since $q_\beta(Y) = 0$ for $\beta \leq 0.9$ and $q_\beta(Y) = 1$ for $\beta > 0.9$, the formula (2) gives that $\mathcal{R}(Y) = 0.5$. The formula (3) gives the same answer. A risk-averse decision maker with this decision model would use 0.5 in comparison with other designs. We now consider the dual expression. In this case, with the scaling $1/(1-\alpha) = 5$, (8) simplifies to

$$\mathcal{R}(Y) = \text{ the maximum value of } 0.9 \cdot 0 \cdot q_1 + 0.1 \cdot 1 \cdot q_2 \text{ such that } 0 \leq q_1, q_2 \leq 5 \text{ and } 0.9q_1 + 0.1q_2 = 1,$$

which has the optimal solution $q_1 = 5/9$ and $q_2 = 5$. The maximum value then becomes $0.9 \cdot 0 \cdot 5/9 + 0.1 \cdot 1 \cdot 5 = 0.5$ that confirms the previous calculation of $\mathcal{R}(Y)$. More interestingly however, the expression can be interpreted as the assessment made by a risk-neutral decision maker that has the nominal distribution with probabilities 0.9 and 0.1 for the realizations 0 and 1, respectively, but that is uncertain about the validity of this distribution. To compensate, she allows the probabilities to be scaled up with a factor of at most 5, while still making sure that they sum to one, in a manner that is the *least* favorable. This risk-neutral decision maker then makes the exact same assessment of the situation as the risk-averse decision maker.

The above dual expression also points to an approach for determining the risk averseness of a specific decision maker. The decision maker could be asked to provide a nominal distribution of costs and then to identify the level of uncertainty associated with that distribution. From these pieces of information a set $\mathcal{Q}$ can be constructed for a family of risk measures that then leads to a specific measure of risk through (8). Moreover, a specific risk envelope indicates the potential benefits derived from improved probabilistic models, which might correspond to narrowing down the risk envelope.

4 Risk Measures in Design Optimization

With the possibility of not only a few but an uncountable number of designs, it becomes essential to enrich decision models with a specific formulation of optimality involving an objective function, to be minimized, and a series of constraints to be satisfied. The resulting canonical optimization formulation, relying on risk measures, takes the form

$$\text{minimize } \mathcal{R}_0(Y_0(\boldsymbol{x})) \text{ subject to } \mathcal{R}_i(Y_i(\boldsymbol{x})) \leq b_i,\ i = 1, 2, ..., I, \text{ and } \boldsymbol{x} \in \mathcal{X},$$

where $\mathcal{R}_i$, $i = 0, 1, ..., I$, are risk measures, possibly different, applied to a collection of response and/or cost random variables $Y_i(\boldsymbol{x})$, $i = 0, 1, ..., I$, each dependent on the design vector $\boldsymbol{x}$. The right-hand sides b_i, $i = 1, 2, .., I$, are constants. As a special case, one may simply have that $Y_i(\boldsymbol{x})$ is a deterministic function $f_i(\boldsymbol{x})$. If $\mathcal{R}_i$ is regular, then $\mathcal{R}_i(Y_i(\boldsymbol{x})) = f_i(\boldsymbol{x})$. Another case is a failure probability constraint $p(Y_i(\boldsymbol{x})) \leq 1 - \alpha$, where, for example, $Y_i(\boldsymbol{x}) = g_i(\boldsymbol{x}, \boldsymbol{V})$, with g_i a limit-state function parameterized by the design variables $\boldsymbol{x}$ and a random vector $\boldsymbol{V}$. In that case, in view of (1), the constraint equivalently takes the form $\mathcal{R}_i(Y_i(\boldsymbol{x})) \leq b_i$, with $\mathcal{R}_i = q_\alpha$ and $b_i = 0$. Consequently, the formulation includes the possibility of minimizing some deterministic cost subject to reliability-type constraints and many other

cases, for examples involving multiple costs and responses simultaneously. The set $\mathcal{X}$ highlights the flexibility by also including constraints that are "simple," i.e., are deterministic and require no special treatment in the implementation. We again stress that the choice of random variable to include in the objective function is somewhat arbitrary as a formulation in terms of a constraint is essentially equivalent.

A key property of regular measures of risk is that the canonical formulation is a convex optimization problem whenever $Y_i(\boldsymbol{x})$, $i = 0, 1, ..., I$, are affine functions of $\boldsymbol{x}$ for every realization, possibly except for an event with probability zero, and $\mathcal{X}$ is a convex set. If $\mathcal{R}_i$ is monotone, then linearity can be relaxed to convexity; see the convexity theorem of Rockafellar and Uryasev (2013). The value of convexity of an optimization problem cannot be overestimated as it dramatically improves the ability of algorithms to obtain globally optimal solutions efficiently. In the absence of convexity, a globally optimal solution is usually inaccessible unless $\boldsymbol{x}$ only involves a small number of variables and a huge computational effort is employed. The result also provides an incentive for constructing linear or convex models (in $\boldsymbol{x}$) for the random variables $Y_i(\boldsymbol{x})$, possibly through approximations. If successful, the benefits for such an effort will be plentiful when solving the canonical formulation.

The trade-off formula (7) allows a simplification of the canonical formulation into the following equivalent form (see the regret theorem of Rockafellar and Uryasev (2013)):

$$\text{minimize } c_0 + \mathcal{V}_0(Y_0(\boldsymbol{x}) - c_0) \text{ subject to } c_i + \mathcal{V}_i(Y_i(\boldsymbol{x}) - c_i) \leq b_i,\ i = 1, 2, ..., I, \text{ and } \boldsymbol{x} \in \mathcal{X},$$

where $\mathcal{V}_i$ are the regular measures of risk corresponding to the regular risk measures $\mathcal{R}_i$ through (7) and $c_0, c_1, ..., c_I$ are auxiliary design variables. This equivalent form is computationally beneficial as the expressions for regret are usually simpler than those for risk. For example, if $\mathcal{R}_i(Y_i(\boldsymbol{x})) = \bar{q}_\alpha(Y_i(\boldsymbol{x}))$, i.e., using a regular and coherent superquantile risk measure, then $\mathcal{V}_i(Y_i(\boldsymbol{x})) = \frac{1}{1-\alpha} E[\max\{0, Y_i(\boldsymbol{x})\}]$ and the constraint $\mathcal{R}_i(Y_i(\boldsymbol{x})) \leq b_i$ takes the following equivalent form

$$c_i + \frac{1}{1-\alpha} E[\max\{0, Y_i(\boldsymbol{x}) - c_i\}] \leq b_i,$$

which simply involves an expectation. If $Y_i(\boldsymbol{x}) = g_i(\boldsymbol{x}, \boldsymbol{V})$ for some function g_i and the distribution of $\boldsymbol{V}$ is discrete with realizations $\boldsymbol{v}^1, ..., \boldsymbol{v}^J$ and probabilities γ^j, then the constraint simplifies further to the collection of constraints

$$\begin{aligned} c_i + \frac{1}{1-\alpha} \sum_{j=1}^{J} \gamma^j a_i^j &\leq b_i, \\ g_i(\boldsymbol{x}, \boldsymbol{v}^j) - c_i &\leq a_i^j, \text{ for all } j = 1, ..., J \\ 0 &\leq a_i^j, \text{ for all } j = 1, ..., J, \end{aligned}$$

with a_i^j, $j = 1, ..., J$, being auxiliary design variables. The conversion of a single constraint into this collection of constraints involving additional variables may at first appear counterproductive, but the simplicity of the reformulation more than outweighs the increase in problem size. In particular, the development of derivative formulae with respect to $\boldsymbol{x}$ for the various constraints is now greatly simplified as it only involves the gradient of g_i with respect to its first argument.

In general, the use of regular risk measures significantly improves our ability to solve the canonical formulation through its reformulations. The deviation from regularity, in contrast, causes difficulties as illustrated by the comparison of a non-regular quantile risk measure, which corresponds to the failure probability, and a regular superquantile risk measure, corresponding to a buffered failure probability. The choice of the quantile risk measure results in a nonconvex optimization problems even if the underlying g_i-functions are linear in the design $\boldsymbol{x}$, the prevalence of locally optimal designs, difficulties with computing gradients with respect to $\boldsymbol{x}$, and temptation to resort to approximations with unknown accuracy. In contrast, a superquantile risk measure leads to convex optimization problems when underlying functions are convex in $\boldsymbol{x}$, globally optimal designs in the absence of other complications, and simple derivative expressions. In fact, the focus on regular measures of risk allows design optimization with little additional complication beyond what is inherent in the response and cost random variables. In contrast, under non-regular measures of risk, the process of design optimization often becomes intractable without a series of additional approximations that could seriously deteriorate the quality of the obtained design.

5 Conclusions

We have described possibilities in risk-averse decision making beyond traditional expected utility theory for cost random variables and failure probability expressions for response random variables. The possibilities center on quantification of risk by *risk measures*, which, in fact, make no distinction between cost and response random variables. Both kinds of random variables could be assessed by any of the resulting decision models. The large number of possible models provides flexibility, but also the need for guidance. In a specific situation, the narrowing down of this vast class is invariably necessary. We give regularity, coherency, and other conditions that limit the choices by focusing on measures of risk that are "reasonable;" superquantile risks emerge as especially convenient measures with decision theoretic foundations in dual utility theory. Still, it is necessary to interview a decision maker, examine her ability to handle high costs and responses, and calibrate with other decisions to quantify exactly the parameters in a selected risk measure. This process is already standard in several areas where utility functions are regularly elicited from decision makers and where limits for probabilities of failure are systematically determined through calibration with existing systems. The connections we make between relative utility and risk measure and between a risk-averse decision maker and a risk-neutral decision maker with distributional uncertain provide further insight that may help in the process of determining a risk measure and its parameters. Although the process of selecting a specific risk measure is undoubtable challenging, our goal with the paper is to provide a solidly founded framework within which that process can be carried out. By adhering to general recommendations such as regularity, significant benefits materialize. One avoids paradoxes that counter our intuition of the meaning of "risk" and one obtains computational tractability and stable solutions. In view of the advantages emerging from these classes of risk measures, we are hopeful that they may lead to improved procedures for risk-averse engineering design and decision making.

Appendix

This appendix summarizes the classical and dual utility theories.

Expected utility theory. The theory was originally developed in terms of probability distributions on the real line (von Neumann and Morgenstern, 1944), but we here follow Ruszczynski (2013) and state axioms in terms of random variables, which helps in the comparison with the dual utility theory.

We need the following notation. For random variables X and Y, we denote by $X \preceq Y$ that X is preferred over Y. We write $X \prec Y$ if $X \preceq Y$, but $Y \preceq X$ does not hold and then say that X is strictly preferred over Y. Indifference between X and Y is denoted by $X \sim Y$ and takes place if $X \preceq Y$ and $Y \preceq X$. We refer to $\preceq$ as a *preference relation*. We always assume that preferences are law-invariant, i.e., if the distribution of X and Y are identical, then $X \sim Y$.

For random variables X, Y, Z, we say that Z is a *lottery* of X and Y with probability α if there is an event with probability α such that the conditional distribution of Z, conditional on that event, is the same as the distribution of X and that the conditional distribution of Z, conditional on the complement of that event, is the same as the distribution of Y. That is, facing the possible outcomes of Z is equivalent to facing the outcomes of X, with probability α, as well as facing the outcomes of Y, with probability $1 - \alpha$. The axioms of utility theory is given in terms of such lotteries:

For any random variables X, Y, either $X \prec Y$ or $Y \prec X$ or $X \sim Y$ (completeness),
If $X \preceq Y$ and $Y \preceq Z$, then $X \preceq Z$ (transitivity),
If $X \prec Y$, then the lottery of X and Z with probability α is strictly preferred to the lottery of Y and Z with probability α, for any α (independence),
If X is strictly preferred over Y and Y strictly preferred over Z, then there exist probabilities α, β such that the lottery of X and Z with α is strictly preferred to Y as well as Y is strictly preferred to the lottery of X and Z with β (archimedean).

The completeness axiom simply states that we are able to express preference or indifference for any pair of random variables in the set of random variables under consideration. The transitivity axiom expresses the natural condition that if a decision maker prefers a first random variable over a second one, and the second one over a third one, then it will also prefer the first random variable over the third. The independence axiom postulates that a preference remains unchanged under the possibility of another outcome. The archimedean axiom expresses that a strict preference is maintained under small changes in the random variables.

A *numerical representation* U of a preference relation $\preceq$ is a functional that for every random variable X under consideration returns a real number $U(X)$ and

$$X \prec Y \text{ if and only if } U(X) > U(Y).$$

That is, a comparison of $U(X)$ and $U(Y)$ suffices to determine a preference of X over Y. The fundamental theorem of von Neumann and Morgenstern (von Neumann and Morgenstern, 1944) states that if a preference relation $\preceq$ satisfies the above axioms, then a numerical representation exists for that

preference relation. Further characterizations are possible under a continuity[‡] axiom:

$$\text{For every } X, \text{ the sets } \{Y : Y \preceq X\} \text{ and } \{Y : X \preceq Y\} \text{ are closed (continuity).} \tag{9}$$

We then have that under the completeness, transitivity, independence, and continuity axioms, there exists a continuous and bounded function u such that

$$U(Y) = E[u(-Y)]$$

is a numerical representation. In this case it is therefore sufficient to rely on expected utility for the characterization of preferences. A monotonicity axiom provides additional details about the nature of u:

$$\text{If } X \leq Y \text{ with probability 1, then } X \preceq Y \text{ (monotonicity).}$$

Under the additional axiom of monotonicity, u is nondecreasing and one can limit the focus to such utility functions.

Dual utility theory. In contrast to the axioms for expected utility theory, which are based on lotteries, the axioms of dual utility theory rely on comparison of *co-monotone random variables*; see for example Ruszczynski (2013) for a recent exposition which we follow here. The definition of co-monotone random variables requires us to view the random variables as (measurable) functions from a common sample space Ω into the real numbers and denote by $X(\omega)$ the realization of random variable X corresponding to outcome $\omega \in \Omega$. Specifically, we say that random variables X and Y are co-monotone if

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geq 0 \text{ for every } \omega, \omega' \in \Omega.$$

In essence, two random variables are co-monotonic if they move in tandem: if one outcome gives a larger realization of X than another outcome, then Y also has larger realizations for the latter outcome. One can view co-monotonicity as a "strong from" of positive correlation. We say that a set of random variables are co-monotone if all the pairs of random variables in the set is co-monotone.

Yaari postulated the following axioms (Yaari, 1987):

$$\begin{aligned}
&\text{For co-monotone random variables } X, Y, Z, \text{ we have that if } X \prec Y,\\
&\quad \text{then } \alpha X + (1-\alpha)Z \prec \alpha Y + (1-\alpha)Z \text{ for all probabilites } \alpha \text{ (dual independence),}\\
&\text{For co-monotone random variables } X, Y, Z, \text{ with } X \prec Y \prec Z, \text{ there exist probabilities}\\
&\quad \alpha, \beta \text{ such that } \alpha X + (1-\alpha)Z \prec Y \prec \beta X + (1-\beta)Z \text{ (dual archimedean).}
\end{aligned}$$

We observe that instead of lotteries, involving alternative outcomes which sometimes cause humans to make "irrational" decisions, the axioms of dual utility theory involve only statements about random

[‡]Closedness in the subsequent statement is with respect to convergence in distribution, i.e., for any sequence Z_n of random variables in a set under consideration that converges in distribution to a random variable Z, we have that Z is in the set.

variables moving in tandem. The dual independence axiom states that if X is strictly preferred over Y, then "mixing" X with a random variable Z and Y with Z, when they all move in tandem, do not change the preference. Since these axioms involve statements only over the co-monotone random variables, they are in some sense weaker than those of expected utility theory.

If the completeness, transitivity, monotonicity, dual independence, and dual archimedean axioms hold for a preference relation, then there exists a numerical representation D characterizing that relation. Further specification of the numerical representation is also possible. If a preference relation is continuous[§] and the consideration is limited to bounded random variables, then there exists a bounded, nondecreasing, and continuous function w such that the numerical representation satisfies (5) for all Y. Consequently, every preference relation over such random variables can be fully captured by a rank-dependent utility function w and (5). We refer to Dentcheva and Ruszczyński (2013) for a modern perspective on expected utility and dual utility theories.

Acknowledgement

The authors thank Prof. T. Haukaas, University of British Columbia, and A. Der Kiureghian, University of California, Berkeley for eye-opening discussions and comments.

References

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

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

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

R.I. Bot, N. Lorenz, and G. Wanka. Dual representations for convex risk measures via conjugate duality. *Journal of Optimization Theory and Applications*, 144:185–203, 2010.

E.J. Cha and B.R. Ellingwood. Risk-averse decision-making for civil infrastructure exposed to low-probability, high-consequence events. *Reliability Engineering and System Safety*, 104(1):27–35, 2012.

E.J. Cha and B.R. Ellingwood. Seismic risk mitigation of building structures: the role of risk aversion. *Structural Safety*, 40(1):11–19, 2013.

C.W. Commander, P.M. Pardalos, V. Ryabchenko, S. Uryasev, and G. Zrazhevsky. The wireless network jamming problem. *Journal of Combinatorial Optimization*, 14:481–498, 2007.

[§]Now closeness in the continuity axiom (9) is taken with respect to convergence in the L_1-norm, i.e., for any sequence Z_n of random variables in a set under consideration with $E[|Z_n - Z|] \to 0$ for a random variable Z, we have that Z is in the set.

F. Delbaen. Coherent measures of risk on general probability spaces. In P.J. Schonbucher K. Sandmann, editor, *Advances in Finance and Stochastics, Essays in Honor of Dieter Sondermann*, pages 1–37. Springer, Berlin, Germany, 2002.

D. Dentcheva and A. Ruszczyński. Common mathematical foundations of expected utility and dual utility theories. *SIAM Journal on Optimization*, 24, 2013.

O. Ditlevsen and H.O. Madsen. *Strutural Reliability Methods*. Wiley, New York, NY, 1996.

K. Dowd. *Measuring market risk*. Wiley, New York, NY, 2. edition, 2005.

M.H. Faber and N.C. Lind. Common rationale for health and life safety management. Paper for Session 3 of the 2012 OECD High Level Risk Forum, 2013.

M.H. Faber, D. Straub, E. Heredia-Zavoni, and R. Montes-Iturrizaga. Risk assessment for structural designcriteria of FPSO systems. part i: Generic models and acceptance criteria. *Marine Structures*, 28(1):120–133, 2013.

H. Föllmer and A. Schied. *Stochastic Finance*. De Gruyter, New York, NY, 2. edition, 2004.

D. Friedman, R.M. Isaac, D. James, and S. Sunder. *Risky Curves. On the empirical failure of expected utility*. Princeton University Press, Routledge, NY, 2014.

P. Gardoni and C. Murphy. Gauging the societal impacts of natural disasters using a capabilities-based approach. *Disasters: The Journal of Disaster Studies, Policy, and Management*, 34(3):619–636, 2010.

P. Krokhmal, M. Zabarankin, and S. Uryasev. Modeling and optimization of risk. *Surveys in Operations Research and Management Science*, 16(2):49–66, 2011.

M. Mahsuli and T. Haukaas. Computer program for multi-model reliability and optimization analysis. *ASCE Journal of Computing in Civil Engineering*, 27(1):8798, 2013.

R. Minguez, E. Castillo, and J.L. Lara. Iterative scenario reduction technique to solve reliability based optimization problems using the buffered failure probability. In G. Deodatis, editor, *Proceedings of ICOSSAR*, 2013.

C. Murphy and P. Gardoni. The role of society in engineering risk analysis: A capabilities-based approach. *Risk Analysis*, 26(4):1073–1083, 2006.

C. Murphy and P. Gardoni. The acceptability and the tolerability of societal risks: A capabilities-based approach. *Science and Engineering Ethics*, 14(1):77–92, 2008.

C. Murphy and P. Gardoni. Design, risk and capabilities. In J. van den Hoven and I. Oosterlaken, editors, *Human Capabilities, Technology, and Design*, pages 173–188. Springer, 2012.

J.S. Nathwani, N.C. Lind, and M.D. Pandey. Affordable safety by choice: the life quality method. IRR, Ontario, Canada: University of Waterloo Earthscan, 1997.

G. Pflug. Subdifferential representation of risk measures. *Mathematical Programming*, 108:339–354, 2007.

J. Quiggin. A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 3(4): 323–343, 1982.

R.T. Rockafellar. Coherent approaches to risk in optimization under uncertainty. *INFORMS, Tutorials in Operations Research*, pages 38–61, 2007.

R.T. Rockafellar and J.O. Royset. On buffered failure probability in design and optimization of structures. *Reliability Engineering & System Safety*, 95:499–510, 2010.

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

R.T. Rockafellar and S. Uryasev. The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science*, 18:33–53, 2013.

A. Ruszczynski. Advances in risk-averse optimization. *INFORMS, Tutorials in Operations Research*, pages 168–190, 2013.

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

M.G. Stewart, B.R. Ellingwood, and J. Mueller. Homeland security: a case study in risk aversion for public decision-making. *International Journal of Risk Assessment and Management*, 15(5/6): 367–386, 2011.

J. von Neumann and O. Morgenstern. *Theorey of games and economic behavior*. Princeton University Press, Princeton, NJ, 1944.

M. E. Yaari. The dual theory of choice under risk. *Econometrica*, 55(1):95–115, 1987.

T.Y. Yang, J. Moehle, B. Stojadinovic, and A. Der Kiureghian. Seismic performance evaluation of facilities: Methodology and implementation. *Journal of Structural Engineering*, 135:1146–1154, 2009.