

RANDOM VARIABLES, MONOTONE RELATIONS AND CONVEX ANALYSIS

R. Tyrrell Rockafellar,¹ Johannes O. Royset²

Abstract

Random variables can be described by their cumulative distribution functions, a class of nondecreasing functions on the real line. Those functions can in turn be identified, after the possible vertical gaps in their graphs are filled in, with maximal monotone relations. Such relations are known to be the subdifferentials of convex functions.

Analysis of these connections yields new insights. The generalized inversion operation between distribution functions and quantile functions corresponds to graphical inversion of monotone relations. In subdifferential terms, it corresponds to passing to conjugate convex functions under the Legendre-Fenchel transform. Among other things, this shows that convergence in distribution for sequences of random variables is equivalent to graphical convergence of the monotone relations and epigraphical convergence of the associated convex functions.

Measures of risk that employ quantiles (VaR) and superquantiles (CVaR), either individually or in mixtures, are illuminated in this way. Formulas for their calculation are seen from a perspective that reveals how they were discovered. The approach leads further to developments in which the superquantiles for a given distribution are interpreted as the quantiles for an overlying “superdistribution.” In this way a generalization of Koenker-Basset error is derived which lays a foundation for superquantile regression as a higher-order extension of quantile regression.

Keywords: *random variables, quantiles, superquantiles, superexpectations, superdistributions, convergence in distribution, stochastic dominance, comonotonicity, measures of risk, value-at-risk, conditional-value-at-risk, convex analysis, conjugate duality, stochastic optimization.*

20 December 2012

¹Mathematics Dept., University of Washington, Seattle, WA 98195-4350; *rtr@uw.edu*

²Operations Research Dept., Naval Postgraduate School, Monterey, CA 93943; *joroyset@nps.edu*

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE 20 DEC 2012	2. REPORT TYPE	3. DATES COVERED 00-00-2012 to 00-00-2012			
4. TITLE AND SUBTITLE Random Variables, Monotone Relations and Convex Analysis		5a. CONTRACT NUMBER			
6. AUTHOR(S)		5b. GRANT NUMBER			
		5c. PROGRAM ELEMENT NUMBER			
		5d. PROJECT NUMBER			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School, Operations Research Dept, Monterey, CA, 93943		5e. TASK NUMBER			
		5f. WORK UNIT NUMBER			
		8. PERFORMING ORGANIZATION REPORT NUMBER			
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)			
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)			
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES in review					
14. ABSTRACT Random variables can be described by their cumulative distribution functions, a class of nondecreasing functions on the real line. Those functions can in turn be identified, after the possible vertical gaps in their graphs are filled in, with maximal monotone relations. Such relations are known to be the subdifferentials of convex functions. Analysis of these connections yields new insights. The generalized inversion operation between distribution functions and quantile functions corresponds to graphical inversion of monotone relations. In subdifferential terms, it corresponds to passing to conjugate convex functions under the Legendre-Fenchel transform. Among other things, this shows that convergence in distribution for sequences of random variables is equivalent to graphical convergence of the monotone relations and epigraphical convergence of the associated convex functions. Measures of risk that employ quantiles (VaR) and superquantiles (CVaR), either individually or in mixtures, are illuminated in this way. Formulas for their calculation are seen from a perspective that reveals how they were discovered. The approach leads further to developments in which the superquantiles for a given distribution are interpreted as the quantiles for an overlying ?superdistribution.? In this way a generalization of Koenker-Basset error is derived which lays a foundation for superquantile regression as a higher-order extension of quantile regression.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 31	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

1 Introduction

The aim of this article is to promote a way of looking at fundamental concepts in probability and statistics by embedding them in a framework of convex analysis. The key is a thorough duality between cumulative distribution functions on $(-\infty, \infty)$ and quantile functions on $(0, 1)$, based on identifying them with the one-sided derivatives of conjugate pairs of convex functions.

Motivation for this framework comes from the modeling of risk in optimization under uncertainty, along with applications to stochastic estimation and approximation. Sharply in focus, beyond distribution functions and quantiles are “superquantiles,” which are quantifications of random variables now recognized as essential building-blocks for “measures of risk” in finance and engineering. Superquantiles fit most simply and naturally with random variables having a cost/loss/damage orientation, in tune with the conventions of optimization theory in which functions are minimized and inequality constraints are normalized to “ \leq ” form. The upper tails of the distributions of such random variables are usually then of more concern than the lower tails. Corresponding adjustments in formulation and terminology from previous work having the opposite orientation is one of our ongoing themes here. First- and second-order stochastic dominance are adapted to this perspective, in particular.

A further benefit of the convex analysis framework is new characterizations of convergence in distribution, a widely used property of approximation. Our analysis indicates moreover how quantile regression, as an alternative to least-squares regression in statistics, can be bootstrapped into a new higher-order approximation tool centered instead on superquantiles. Helpful estimates of superquantiles, for numerical work and more, are derived as well. Second-derivative duality in convex analysis further produces a duality between distribution densities and quantile densities.

Distribution functions versus quantile functions. The path to these developments begins with elementary observations in a two-dimensional graphical setting with pairs of non-decreasing functions in an extended inverse-like relationship.

A real-valued random variable X gives a probability measure on the real line \mathbb{R} which can be described by the (cumulative) *distribution function* F_X for X , namely

$$F_X(x) = \text{prob}\{X \leq x\} \quad \text{for } x \in (-\infty, \infty). \quad (1.1)$$

The function F_X is nondecreasing and right-continuous on $(-\infty, \infty)$, and it tends to 0 as $x \rightarrow -\infty$ and to 1 as $x \rightarrow \infty$. These properties characterize the class of functions that furnish distributions of random variables. Right-continuity of a nondecreasing function reduces to continuity except at points where the graph has a vertical gap. The set of such jump points, if any, has to be finite or countably infinite.

The probability measure associated with a random variable X can alternatively be described by its *quantile function* Q_X , namely

$$Q_X(p) = \min\{x \mid F_X(x) \geq p\} \quad \text{for } p \in (0, 1), \quad (1.2)$$

so that $Q_X(p)$ is the lowest x such that $\text{prob}\{X > x\} \leq 1-p$. The function Q_X is nondecreasing and left-continuous on $(0, 1)$, and those properties characterize the class of functions that furnish

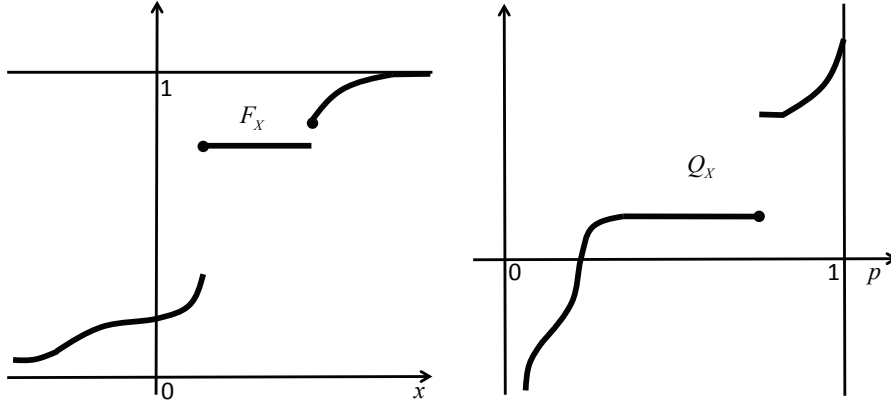


Figure 1: distribution function F_X and quantile function Q_X

quantiles of random variables. The correspondence between distribution functions and quantile functions is one-to-one, with F_X recoverable from Q_X by

$$F_X(x) = \begin{cases} \max\{p \mid Q_X(p) \leq x\} & \text{for } x \in (\inf Q_X, \sup Q_X], \\ 1 & \text{for } x > \sup Q_X, \\ 0 & \text{for } x \leq \inf Q_X. \end{cases} \quad (1.3)$$

The quantile function Q_X , like the distribution function F_X , can have at most countably many jumps where it fails to be continuous. The vertical gaps in the graph of Q_X correspond to the horizontal segments in the graph of F_X , and vice versa, as seen in Figure 1. It follows that F_X and Q_X can likewise have at most countably many horizontal segments in their graphs.

When the graph of F_X has no vertical gaps or horizontal segments, so that F_X is not only continuous but (strictly) increasing, the “min” in (1.2) is superfluous and $Q_X(p)$ is the unique solution x to $F_X(x) = p$. Then Q_X is just the inverse F_X^{-1} of F_X on $(0, 1)$. Without such restriction, though, one can only count on $Q_X(F_X(x)) \geq x$ and $F_X(Q_X(p)) \leq p$, along with

$$F_X(x) \geq p \iff Q_X(p) \leq x. \quad (1.4)$$

The generalized inversion represented in (1.4) and formulas (1.2) and (1.3) can be given a solid footing in geometry. By filling in the vertical gaps in the graphs of F_X and Q_X , and adding infinite vertical segments at the right and left ends of the resulting “curve” for Q_X to mimic the infinite horizontal segments which appear at the ends of the graph of F_X when the

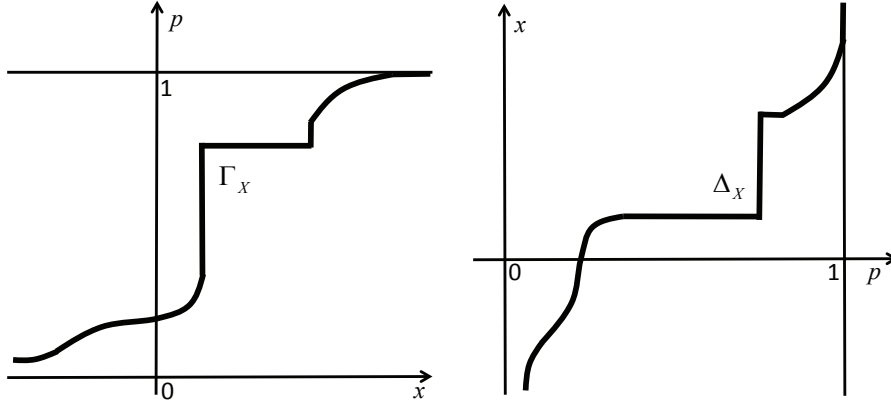


Figure 2: the relation Γ_X and its inverse Δ_X

range of X is bounded from above or below, one obtains the “curves” Γ_X and Δ_X of Figure 2. These “curves” are the reflections of each other across the 45-degree line in $\mathbb{R} \times \mathbb{R}$ where $x = p$.

In the classical mindset it would be anathema to fill in vertical gaps in the graph of a function, thereby ruining its status as a “function” (single-valued). In this situation, though, there are overriding advantages. The graphs Γ_X and Δ_X belong to a class of subsets of \mathbb{R}^2 called *maximal monotone relations*. Such relations have powerful properties and are basic to convex analysis, which identifies them with the “subdifferentials” of convex functions. This will be recalled in Section 2. The graphical inversion in Figure 2, where

$$\Delta_X = \{ (p, x) \mid (x, p) \in \Gamma_X \}, \quad \Gamma_X = \{ (x, p) \mid (p, x) \in \Delta_X \},$$

will be portrayed there as corresponding to the Legendre-Fenchel transform, which dualizes a convex function by pairing it with a *conjugate* convex function.

Although monotone relations are central in this paper, the idea of looking at conjugate pairs of convex functions defined in one way or another through direct integration of F_X and Q_X is not new, cf. Ogryczak and Ruszczyński [14] and subsequently [15], [16]. What is different here is a choice of functions that better suits random variables with cost/loss/damage orientation in handling their upper tails. The need of such a switch for purposes in stochastic optimization has recently motivated Dentcheva and Martinez [4] to adapt also in that direction, but our approach seems to achieve that more simply and comprehensively.

The convergence theory for maximal monotone relations and the convex functions having

them as their subdifferentials can be coordinated in our framework with results about the convergence of sequences of random variables. A special feature is that approximations on the side of convex analysis are most effectively studied through “set convergence.” Maximal monotone relations are compared in terms of distances to their graphs, while convex functions are compared in terms of distances to their epigraphs rather than their graphs. Indeed, epigraphical convergence of convex functions is tantamount to graphical convergence of their subdifferentials. Epigraphical convergence is the only topological convergence that renders the Legendre-Fenchel transform continuous.

For a sequence of random variables X_k this leads, as we demonstrate in Section 3, to fresh characterizations of “convergence in distribution” of X_k to a random variable X . It corresponds to graphical convergence of the associated maximal monotone relations Γ_{X_k} to Γ_X , or for that matter Δ_{X_k} to Δ_X , and to epigraphical convergence of the convex functions having their subdifferentials described by those relations. That epigraphical convergence, in this special context, can essentially be reduced to pointwise convergence.

Superquantile functions. In associating with Γ_X a convex function having it as subdifferential, and then investigating the conjugate of that convex function, information is gained about *superquantiles* of random variables. Superquantiles refer to values which, like quantiles, capture all the information about the distribution of a random variable, but in doing that avoid some of the troublesome properties of quantiles such as potential discontinuity and instability with respect to parameterization. They have been studied under different names for modeling risk in finance, but here we are translating them to the general theory of statistics and probability. Bringing out their significance in that environment is one of our goals.

For a random variable X with cost/loss/damage orientation, the superquantile $\bar{Q}_X(p)$ at probability level $p \in (0, 1)$ has two equivalent expressions which look quite different. First,

$$\bar{Q}_X(p) = \text{expectation in the (upper) } p\text{-tail distribution of } X. \quad (1.5)$$

This refers to the probability distribution on $[Q_X(p), \infty)$ which, in the case of $F_X(Q_X(p)) = p$, is the conditional distribution of X subject $X \geq Q_X(p)$, but which “rectifies” that conditional distribution when F_X has a jump at the quantile $Q_X(p)$, so that $F_X(Q_X(p)) > p$. In the latter case there is a probability atom at $Q_X(p)$ causing the interval $[Q_X(p), \infty)$ to have probability larger than $1 - p$ and the interval $(Q_X(p), \infty)$ to have probability smaller than $1 - p$. To take care of the discrepancy, the p -tail distribution is defined in general as having $F_X^{[p]}(x) = \max\{0, F_X(x) - p\}/(1 - p)$ as its distribution. This amounts to an appropriate splitting of the probability atom at $Q_X(p)$. The second expression for the superquantile is

$$\bar{Q}_X(p) = \frac{1}{1 - p} \int_p^1 Q_X(p') dp'. \quad (1.6)$$

The equivalence between the two expressions will be explained in Section 3, which will also clarify the restrictions that need to be imposed to ensure both are well defined.

In finance, the quantile $Q_X(p)$ is identical to the popular notion of the *value-at-risk* $\text{VaR}_p(X)$ of X at probability level p .³ The superquantile $\bar{Q}_X(p)$ as defined by (1.5) goes back to Rockafellar and Uryasev [22] (and an earlier working paper of 1999), with follow-up in [23]. There it was

³This is the case for cost/loss-oriented random variables. In applications centered on random variables Y

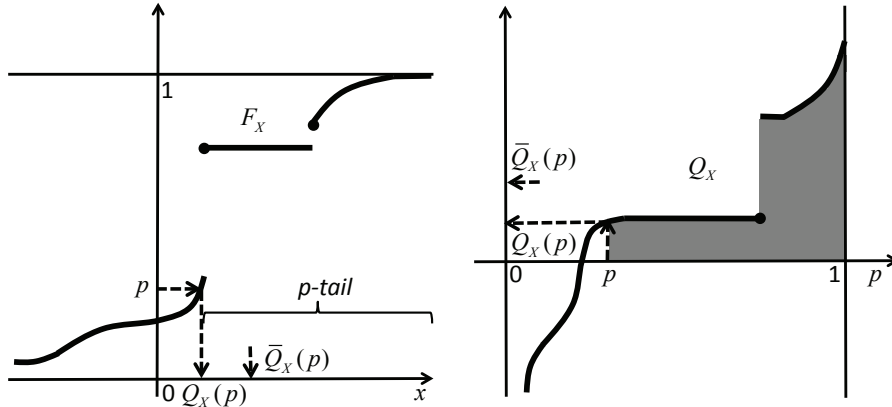


Figure 3: the p th quantile and the p -tail

called *conditional-value-at-risk*, as suggested by (1.5) and its interpretation as the conditional expectation of X subject to $X \geq Q_X(p)$ when F_X has no jump at the quantile $Q_X(p)$. It was denoted by $\text{CVaR}_p(X)$ in order to contrast it with $\text{VaR}_p(X)$. The expression on the right side of (1.6) was independently introduced around the same time by Acerbi [1] as “expected shortfall,”⁴ but the equivalence of the two was soon realized. Because statistical terminology ought to be free from dependence on financial terminology, we think it helpful to have “superquantile” available as a neutral alternative name. This was suggested in our paper [20] on reliability in engineering and has been pursued further in the “risk quadrangle” setting of [24].

That side-by-side approach advantageously suggests making a graphical comparison between the superquantile function \bar{Q}_X and the quantile function Q_X , as in Figure 4. A new and immediate insight is that \bar{Q}_X is the inverse of a distribution function \bar{F}_X generated from F_X . We call this the corresponding *superdistribution* function. It lets the superquantiles of X be identified as the quantiles for an auxiliary probability measure on $(-\infty, \infty)$. Specifically, \bar{F}_X is the distribution function for an auxiliary random variable \bar{X} derived from X , and this will

that are profit/gain-oriented, the value-at-risk of Y at probability level p is $-Q_{[-Y]}(1-p)$. The avoidance of such complications with minus signs is one of the reasons why we prefer cost/loss orientation in setting forth principles for use in statistics and probability with applications to optimization.

⁴The interpretation of this integral as an average led Föllmer and Schied [7] to instead call this quantity “average value-at-risk” with notation AVaR.

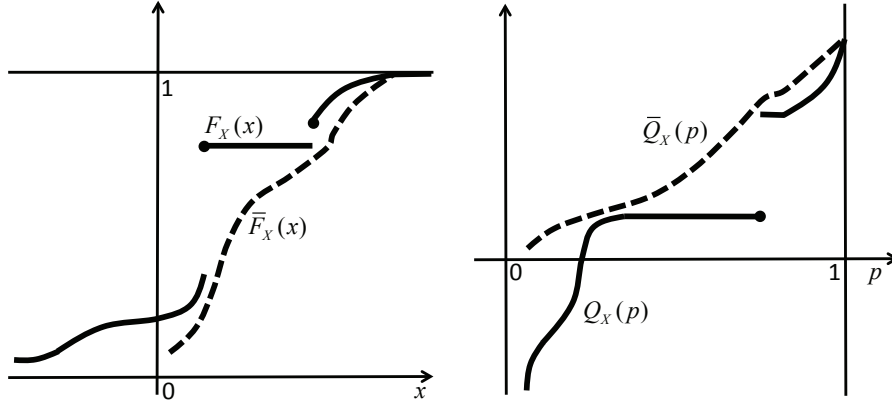


Figure 4: superdistribution function \bar{F}_X and superquantile function \bar{Q}_X

have valuable consequences:

$$\bar{F}_X = F_{\bar{X}} \text{ for the random variable } \bar{X} = \bar{Q}_X(F_X(X)). \quad (1.7)$$

Motivating connections with risk. Although superquantiles have not previously been touted as a potentially significant addition to basic statistics, their importance in formulating problems of stochastic optimization is already well recognized. A brief discussion of “risk” will help in understanding the interest in them coming from that direction.

A *measure of risk* is a functional \mathcal{R} that assigns to a random variable X a value $\mathcal{R}(X)$ in $(-\infty, \infty]$ as a quantification of the risk in it.⁵ The context here is that of X representing a generalized “cost,” “loss” or “damage” index, meaning that lower values are preferred to higher values. Typically it is desired to have the outcomes of X below a threshold b , but some violations may have to be accepted. For instance, it would be nice if the losses for a given portfolio of financial assets were always ≤ 0 , but arranging for that might not be feasible. How then can trade-off preferences be captured? How can the desire to have X be “adequately” $\leq b$ in its outcomes be given a mathematical formulation? The role of a risk measure \mathcal{R} is to model this as $\mathcal{R}(X) \leq b$.

Specific examples can help in appreciating the issues. In taking $\mathcal{R}(X) = E[X]$ (expectation), the interpretation of $\mathcal{R}(X) \leq b$ is that the outcomes of X are $\leq b$ “on average.” That choice

⁵Measures of risk are not “measures” in the usual sense of mathematics. This terminology, in which a “measure” is a “quantification,” is widespread in finance.

could be strengthened by taking the measure of risk to be $\mathcal{R}(X) = E[X] + \lambda\sigma(X)$ for a parameter value $\lambda > 0$, where $\sigma(X)$ denotes standard deviation. Then the interpretation of $\mathcal{R}(X) \leq b$ is that outcomes of X above b can only be in the part of the distribution of X lying more than λ standard deviation units beyond the expectation. Such a “safety margin” approach is attractive for its resemblance to confidence levels in statistics. A third choice of risk measure, aimed at enforcing certainty, is to take $\mathcal{R}(X) = \sup X$ (the essential supremum of X , which might be ∞). Then $\mathcal{R}(X) \leq b$ means there is zero probability of an outcome $> b$.

Two further examples, more directly in line with our interests in this article, are quantiles, $\mathcal{R}(X) = Q_X(p)$, and superquantiles, $\mathcal{R}(X) = \overline{Q}_X(p)$, as measures of risk. The corresponding interpretations of having $\mathcal{R}(X) \leq b$ are as follows:

$$Q_X(p) \leq b \iff \text{prob}\{X \leq b\} \geq p \iff \text{prob}\{X > b\} \leq 1 - p, \quad (1.8)$$

$$\overline{Q}_X(p) \leq b \iff \text{even in its } p\text{-tail distribution, } X \text{ is } \leq b \text{ on average.} \quad (1.9)$$

Probabilistic constraints as in (1.8) have a wide following. In contrast, the condition in (1.9) might appear arbitrary and hard to work with. But it has serious motivation in the theory of risk, plus the virtue of taking into account some degree of effects in the upper tail of the distribution of X beyond the threshold b .⁶

A feature of risk theory that elevates superquantiles above quantiles is found in the notion of *coherency* proposed by Artzner et al. [2], originally for purposes of determining appropriate cash reserves in the banking industry. Coherency of a risk measure \mathcal{R} entails having

$$\begin{aligned} \mathcal{R}(C) &= C \quad \text{for constant random variables } X \equiv C, \\ \mathcal{R}(X) &\leq \mathcal{R}(X') \quad \text{when } X \leq X' \text{ almost surely,} \\ \mathcal{R}(X + X') &\leq \mathcal{R}(X) + \mathcal{R}(X'), \\ \mathcal{R}(\lambda X) &= \lambda X \quad \text{for } \lambda > 0. \end{aligned} \quad (1.10)$$

Along with the surface meaning of these axioms,⁷ there are crucial implications for preserving convexity when measures of risk are employed in optimization. This is explained from several angles in [24].

For the examples above, coherency holds for the extreme choices $\mathcal{R}(X) = E[X]$ and $\mathcal{R}(X) = \sup X$, but it is absent in general for $\mathcal{R}(X) = E[X] + \lambda\sigma(X)$ with $\lambda > 0$ (because the monotonicity axiom fails) and for $\mathcal{R}(X) = Q_X(p)$ (because the subadditivity axiom fails). However, coherency does hold for $\mathcal{R}(X) = \overline{Q}_X(p)$. Moreover it holds for weighted sums like $\mathcal{R}(X) = \sum_{k=1}^m \lambda_k \overline{Q}_X(p_k)$ with $\lambda_k > 0$ and $\sum_{k=1}^m \lambda_k = 1$, and even for “continuous” versions of those sums, $\mathcal{R}(X) = \int_0^1 \overline{Q}_X(p) d\lambda(p)$ for a probability measure λ on $(0, 1)$.⁸ In fact, a functional

⁶In reliability terms, with outcomes $X > b$ signaling “failure,” $1 - p$ is the probability of failure in (1.8) and the *buffered* probability of failure in (1.9), cf. [20].

⁷The subadditivity inequality, for instance says, in the context of cash reserves in finance, that if the cash amount $\mathcal{R}(X)$ is adequate for covering the risks in a portfolio with losses described by the random variable X , and $\mathcal{R}(X')$ is enough for a separate portfolio with losses described by X' , then the sum of these amounts should cover the combined portfolio. This supports the idea of diversification of assets. See Föllmer and Schied [7] for more about the role of coherency in finance.

⁸Such expressions relate strongly to “dual utility theory,” the foundations of which have recently been strengthened by Dentcheva and Ruszczyński [6].

$\mathcal{R}(X)$ expressible as the supremum of a collection of such superquantile integrals is known to be the most general kind of coherent measure of risk that depends only on F_X and possesses a certain continuity property; see [10] and [7, Section 2.6].

This makes clear that superquantiles are basic to the foundations of risk theory and further explains why we are intent here on positioning them prominently in view.

Although the defining formulas for superquantiles might raise a perception of them being troublingly complicated or even intractable in comparison to quantiles, quite the opposite is true. A double formula due to Rockafellar and Uryasev [22, 23] brings them together in a way that supports practical methods of computation while bypassing technical issues in the defining formulas (1.5) and (1.6):

$$\begin{aligned}\bar{Q}_X(p) &= \min_x \{ x + \mathcal{V}_p(X - x) \}, & \text{where } \mathcal{V}_p(X) &= \frac{1}{1-p} E[\max\{0, X\}]. \\ Q_X(p) &= \operatorname{argmin}_x \{ x + \mathcal{V}_p(X - x) \} \text{ (left endpoint, if this is not a singleton),}\end{aligned}\tag{1.11}$$

The “argmin,” consisting of the x values for which the minimum is attained, is, in this formula, a nonempty, closed, bounded interval which typically reduces to a single x . The functional \mathcal{V}_p satisfies

$$\begin{aligned}\mathcal{V}_p(X) &\leq \mathcal{V}_p(X') \text{ when } X \leq X' \text{ almost surely,} \\ \mathcal{V}_p(X + X') &\leq \mathcal{V}_p(X) + \mathcal{V}_p(X'), \\ \mathcal{V}_p(\lambda X) &= \lambda \mathcal{V}_p(X) \text{ for } \lambda > 0, \\ \mathcal{V}_p(X) &\geq E[X], \text{ with equality holding only when } X \equiv 0.\end{aligned}\tag{1.12}$$

Such properties are associated with regular “measures of regret” (rather than risk) in the terminology of [24], and it is appropriate therefore, in view of (1.11) to refer to \mathcal{V}_p as *quantile regret*. The functional $\mathcal{E}_p(X) = \mathcal{V}_p(X) - E[X]$ paired with \mathcal{V}_p by [24] as its associated “measure of error” is normalized Koenker-Basset error. It underlies quantile regression as a statistical methodology offering an alternative to least-squares regression [8], [9].

In models of stochastic optimization that incorporate superquantiles in constraints or objectives, the superquantile formula in (1.11) can be substituted in each instance with an associated auxiliary variable in the overall minimization. This greatly simplifies computations and simultaneously yields values for the corresponding quantiles in the solution; cf. [22, 23]. No such computational help is available for constraints and objectives expressed in quantiles instead of superquantiles. As the formulas in (1.11) underscore for anyone familiar with the relative behavior of “min” and “argmin” in numerical optimization, quantiles are inherently less stable than superquantiles in circumstances where random variables depend on decision parameters.

A byproduct of the connections explored here between distributions, monotone relations, and the convex functions associated with them subdifferentially, will be an explanation—for the first time—of how (1.11) was discovered in the background of [22, 23]. It came from recognition of the consequences of applying the Legendre-Fenchel transformation to those convex functions.

A further goal of this article is to develop, in Section 4, a formula along the lines of (1.11) in which the argmin gives the superquantile instead of the quantile:

$$\bar{Q}_X(p) = \operatorname{argmin}_x \{ x + \bar{\mathcal{V}}_p(X - x) \}$$

for the right choice of a “regret” functional $\bar{\mathcal{V}}_p$. Such a formula is needed to for the purpose of generalizing “quantile regression” to “superquantile regression” in the framework of [25] and

[24]. Specifically, from $\bar{\mathcal{V}}_p(X)$ as “superquantile regret,” the functional $\bar{\mathcal{E}}_p(X) = \bar{\mathcal{V}}_p(X) - E[X]$ will be the right substitute for Koenker-Basset error in that generalization. Expressions for superquantile regret and superquantile error that serve in this manner have not previously been identified.

2 Monotone Relations in Convex Analysis

In this section we review facts about monotone relations and the convex functions associated with them in order to lay a foundation for analyzing the connections indicated in Section 1. In that analysis, carried out in Section 3, the x variable will have a quantile role and the p variable will be associated with probability, but for now both are abstract variables with roles completely interchangeable.

Definition (monotonicity and maximal monotonicity). *A set Γ of pairs $(x, p) \in \mathbb{R} \times \mathbb{R}$ is said to give a monotone relation if*

$$(x_1 - x_2)(p_1 - p_2) \geq 0 \quad \text{for all } (x_1, p_1) \text{ and } (x_2, p_2) \text{ in } \Gamma, \quad (2.1)$$

so that either $(x_1, p_1) \leq (x_2, p_2)$ or $(x_1, p_1) \geq (x_2, p_2)$ in the usual coordinatewise ordering of vectors in $\mathbb{R} \times \mathbb{R}$. In other words, a monotone relation is a subset of $\mathbb{R} \times \mathbb{R}$ that is totally ordered in that partial ordering. A monotone relation Γ is maximal if it cannot be enlarged without destroying the total ordering; there is no monotone relation $\Gamma' \supset \Gamma$ with $\Gamma' \neq \Gamma$.

Any monotone relation can be extended to a maximal monotone relation (not necessarily in only one way). Maximal monotonicity was introduced in 1960 by Minty [13] in the study of relations between variables like current and voltage in electrical networks and their analogs in other kinds of networks.

The symmetry in the roles of the two variables in monotonicity has the consequence that if Γ is a monotone relation, then the *inverse* relation Γ^{-1} , defined by

$$\Gamma^{-1} = \{ (p, x) \mid (x, p) \in \Gamma \}, \quad (2.2)$$

is likewise monotone. Maximality passes over in this manner as well.

A maximal monotone relation has the graphical appearance of an unbounded continuous curve that “trends from southwest to northeast” and may incorporate horizontal and vertical segments. It may even begin or end with such a segment of infinite length. As extreme cases, an entire horizontal line gives a maximal monotone relation and so does an entire vertical line. The union of the nonnegative x -axis with the nonpositive p -axis is likewise a maximal monotone relation, moreover one which very commonly arises in applications (not tied to probability). It is the “infinite gamma” shape of that relation that earlier suggested the notation Γ .

A noteworthy feature of a maximal monotone relation is its canonical parameterization by an auxiliary variable t :

$$\begin{aligned} &\text{For a maximal monotone relation } \Gamma \text{ and any } t \in (-\infty, \infty), \\ &\text{the line } x + p = t \text{ intersects } \Gamma \text{ in a unique point } (x(t), p(t)), \\ &\text{and } x(t) \text{ and } p(t) \text{ are Lipschitz continuous as functions of } t. \end{aligned} \quad (2.3)$$

Put in another way, the graph of a maximal monotone relation is a sort of manifold that is globally “lipeomorphic” to the real line. This striking property, so function-like, makes up for the disadvantage, to classical eyes, of allowing the graph to contain vertical segments.

An exposition of the theory of maximal monotone relations which covers (2.3) and other properties yet to be mentioned is available in [26, Chapter 12], where the subject is extended beyond subsets of $\mathbb{R} \times \mathbb{R}$ to subsets of $\mathbb{R}^n \times \mathbb{R}^n$. (In the higher dimensional setting, monotonicity becomes a generalization of positive-semidefiniteness.) Some aspects are also in the earlier book [17, Section 24]. A version of the subject dedicated to extending the original network ideas of Minty, and offering many examples, is in [19, Chapter 8].

Some basic convexity properties are obvious from the Minty parameterization. For instance, the domain and range of a maximal monotone relation Γ , namely

$$\text{dom } \Gamma = \{x \mid (x, p) \in \Gamma \text{ for some } p\}, \quad \text{rge } \Gamma = \{p \mid (x, p) \in \Gamma \text{ for some } x\}, \quad (2.4)$$

are nonempty intervals, although not necessarily closed, while the sets

$$\Gamma(x) = \{p \mid (x, p) \in \Gamma\}, \quad \Gamma^{-1}(p) = \{x \mid (x, p) \in \Gamma\}, \quad (2.5)$$

are closed intervals with $\Gamma(x) \neq \emptyset$ when $x \in \text{dom } \Gamma$, and $\Gamma^{-1}(p) \neq \emptyset$ when $p \in \text{rge } \Gamma$. Clearly $\text{dom } \Gamma^{-1} = \text{rge } \Gamma$ and $\text{rge } \Gamma^{-1} = \text{dom } \Gamma$.

The connection between maximal monotone relations Γ and nondecreasing functions γ on $(-\infty, \infty)$ is elementary and closely reflects the special case of distribution functions considered in Section 1. Suppose $\gamma : (-\infty, \infty) \rightarrow [-\infty, \infty]$ is nondecreasing and not identically $-\infty$ or identically ∞ . Then there are left and right limits

$$\gamma^-(x) = \lim_{x' \nearrow x} \gamma(x'), \quad \gamma^+(x) = \lim_{x' \searrow x} \gamma(x'), \quad (2.6)$$

with $\gamma^-(x) \leq \gamma(x) \leq \gamma^+(x)$. They define functions γ^- and γ^+ which are left-continuous and right-continuous, respectively. A maximal monotone relation Γ is obtained by taking

$$\Gamma = \{(x, p) \in \mathbb{R} \times \mathbb{R} \mid \gamma^-(x) \leq p \leq \gamma^+(x)\}. \quad (2.7)$$

The original γ has no direct role in this and could be replaced by either γ^+ or γ^- from the start, because $(\gamma^+)^- = \gamma^-$ and $(\gamma^+)^+ = \gamma^+$, whereas $(\gamma^-)^- = \gamma^-$ and $(\gamma^-)^+ = \gamma^+$. Conversely, given a maximal monotone relation Γ one can define

$$\begin{aligned} \gamma^-(x) &= \min\{p \mid (x, p) \in \Gamma\} \text{ and } \gamma^+(x) = \max\{p \mid (x, p) \in \Gamma\} \text{ for } x \in \text{dom } \Gamma, \\ \gamma^-(x) &= \gamma^+(x) = -\infty \text{ at points } x \text{ to the left of } \text{dom } \Gamma \text{ (if any),} \\ \gamma^-(x) &= \gamma^+(x) = \infty \text{ at points } x \text{ to the right of } \text{dom } \Gamma \text{ (if any),} \end{aligned} \quad (2.8)$$

to get a pair of nondecreasing functions γ^- and γ^+ , one continuous from the left and one continuous from the right, which produce Γ through (2.7).

Subdifferentiation. The connection between maximal monotone relations and the subdifferentiation of convex functions will be explained next. A *proper convex function* on $(-\infty, \infty)$ is a function $f : (-\infty, \infty) \rightarrow (-\infty, \infty]$ that is not $\equiv \infty$ and satisfies

$$f((1 - \tau)x + \tau x') \leq (1 - \tau)f(x) + \tau f(x') \text{ for all } \tau \in (0, 1) \text{ and all } x, x'. \quad (2.9)$$

In terms of the effective domain and epigraph of f , defined by

$$\text{dom } f = \{x \mid f(x) < \infty\}, \quad \text{epi } f = \{(x, v) \mid f(x) \leq v < \infty\}, \quad (2.10)$$

the definition is equivalent to saying that the proper convex functions are the functions $f : (-\infty, \infty) \rightarrow (-\infty, \infty]$ for which $\text{epi } f$ is a nonempty convex subset of $\mathbb{R} \times \mathbb{R}$, or equivalent on the other hand to taking a nonempty interval I , a finite convex function f on I , and then defining $f(x) = \infty$ for $x \notin I$; then $\text{dom } f = I$.

A proper convex function f is said to be *closed* when it is lower semicontinuous, i.e., has its lower level sets $\{x \mid f(x) \leq c\}$ closed, for all $c \in \mathbb{R}$. This holds if and only if $\text{epi } f$ is closed in $\mathbb{R} \times \mathbb{R}$. Because a finite convex function on an open interval is necessarily continuous, a proper convex function has to be continuous except perhaps at the endpoints of $\text{dom } f$. Closedness thus refers only to behavior at those endpoints. It requires that $f(x_k) \rightarrow f(x)$ when x is an endpoint and a sequence of points $x_k \neq x$ in $\text{dom } f$ tends to x .

For a proper convex function f and any $x \in \text{dom } f$, the left-derivative and the right-derivatives of f , namely,

$$f'^-(x) = \lim_{x' \nearrow x} \frac{f(x') - f(x)}{x' - x}, \quad f'^+(x) = \lim_{x' \searrow x} \frac{f(x') - f(x)}{x' - x}, \quad (2.11)$$

exist with $f'^-(x) \leq f'^+(x)$. The “set-valued” mapping ∂f defined by

$$\partial f(x) = \begin{cases} \{p \in \mathbb{R} \mid f'^-(x) \leq p \leq f'^+(x)\} & \text{for } x \in \text{dom } f, \\ \emptyset & \text{for } x \notin \text{dom } f, \end{cases} \quad (2.12)$$

is called the *subdifferential* of f . When $f'^-(x) = f'^+(x)$, this common value (if finite) is the derivative $f'(x)$. That holds for all but countably many points in the interior of $\text{dom } f$ because of the convexity of f : The one-sided derivatives, as functions of x , are nondecreasing and respectively left-continuous and right-continuous, having $(f'^-)^+ = f'^+$ and $(f'^+)^- = f'^-$.

The key fact about subdifferentials ∂f in general, going beyond the case of single-valuedness, is this:

for a proper convex function f that is closed, the graph of ∂f , namely

$$\text{gph } \partial f = \{(x, p) \mid p \in \partial f(x)\},$$

is a maximal monotone relation Γ ; moreover every maximal monotone relation Γ is the graph of ∂f for some closed proper convex function f , and such a function f is uniquely determined up to an additive constant. (2.13)

The first part of this statement stems from the observation that if the one-sided derivatives in (2.12) are extended outside of $\text{dom } f$ by taking

$$\begin{aligned} f'^-(x) &= f'^+(x) = -\infty && \text{at points } x \text{ to the left of } \text{dom } f \text{ (if any),} \\ f'^-(x) &= f'^+(x) = \infty && \text{at points } x \text{ to the right of } \text{dom } f \text{ (if any),} \end{aligned} \quad (2.14)$$

one gets as $\gamma^- = f'^-$ and $\gamma^+ = f'^+$ a left-continuous/right-continuous pair on all of $(-\infty, \infty)$ for which the Γ associated by (2.7) is $\text{gph } \partial f$. The second part is established by taking for

a given maximal monotone relation Γ the corresponding left-continuous/right-continuous pair γ^-, γ^+ , and for any γ between them and any $x_0 \in \text{dom } \Gamma$ defining

$$f(x) = \int_{x_0}^x \gamma(t)dt + c \quad \text{for an arbitrary constant } c. \quad (2.15)$$

It turns out then that f is a closed proper convex function having $\gamma^- = f'^-$ and $\gamma^+ = f'^+$.

Dualization through the Legendre-Fenchel transform. Duality is the topic reviewed next. Its central feature is a one-to-one correspondence among closed proper convex functions which unites them in pairs. Here we only look at it in the one-dimensional setting, but in multi-dimensional and even infinite-dimensional convex analysis it is the repository of virtually every duality property that is known; see [17, 18, 26].

For any closed proper convex function f on $(-\infty, \infty)$ the formula

$$f^*(p) = \sup_x \{ xp - f(x) \} \quad (2.16)$$

defines a closed proper convex function f^* on $(-\infty, \infty)$ such that

$$f(x) = \sup_p \{ xp - f^*(p) \} = (f^*)^*(x).$$

The passage from f to f^* is the *Legendre-Fenchel transform*. Accompanying these formulas are the subdifferential rules that

$$\partial f^*(p) = \text{argmax}_x \{ xp - f(x) \}, \quad \partial f(x) = \text{argmax}_p \{ px - f^*(p) \}. \quad (2.17)$$

A major consequence for purposes here is that

$$\begin{aligned} &\text{For any conjugate pair of closed proper convex functions, } f \text{ and } f^*, \\ &\text{one has } \partial f^* = (\partial f)^{-1} \text{ meaning that } x \in \partial f^*(p) \iff p \in \partial f(x). \end{aligned} \quad (2.18)$$

Thus, *the maximal monotone relations associated with f and f^* are inverse to each other*. Note that as special cases of (2.17) and (2.18) one has

$$\inf f = -f^*(0), \quad \text{argmin } f = \partial f^*(0), \quad \inf f^* = -f(0), \quad \text{argmin } f^* = \partial f(0). \quad (2.19)$$

Set convergence and its variants. Finally, notions of convergence that are natural to convex analysis need to be explained, particularly because they hardly enter the standard frame of analysis (although they should). We keep to the context of \mathbb{R}^2 because that is all we require, but a full theory in finite dimensions is provided in [26, Chapter 4].

For a nonempty closed subset S of \mathbb{R}^2 , the associated *distance function* is

$$d_S(u) = \min_{w \in S} \|u - w\| \quad \text{for the Euclidean norm } \|\cdot\|.$$

(Any norm would do equally well.) This function d_S is nonnegative with $S = \{u \mid d_S(u) = 0\}$ and it is Lipschitz continuous with Lipschitz constant 1. We are concerned with a sequence of nonempty closed subsets S_k in \mathbb{R}^2 and the issue of whether S_k “converges” to S as $k \rightarrow \infty$, with *set-convergence* in the Kuratowski/Painlevé sense intended. Although there are numerous

characterizations of such convergence (cf. [26, Chapter 4]), it suffices here to concentrate on a description that is easy to visualize:

$$\lim_{k \rightarrow \infty} S_k = S \iff \lim_{k \rightarrow \infty} d_{S_k}(u) = d_S(u) \text{ for every } u.$$

Because of the Lipschitz continuity, such pointwise convergence entails uniform convergence of the distance functions on all bounded sets.

Two convergence notions more closely tuned to our present discussion are built around such set convergence. First,

$$\textit{graphical convergence to } \Gamma \text{ of a sequence of maximal monotone relations } \Gamma_k \text{ refers to their convergence as subsets of } \mathbb{R} \times \mathbb{R}. \quad (2.20)$$

Second,

$$\textit{epigraphical convergence to } f \text{ of a sequence of closed proper convex functions } f_k \text{ on } (-\infty, \infty) \text{ refers to the set-convergence of } \text{epi } f_k \text{ to } \text{epi } f \text{ in } \mathbb{R} \times \mathbb{R}. \quad (2.21)$$

Two celebrated results about these notions underscore their fundamental significance.

$$\text{A sequence of closed proper convex functions } f_k \text{ epi-converges to } f \text{ if and only if the maximal monotone relations } \Gamma_k = \text{gph } \partial f_k \text{ converge graphically to } \Gamma = \text{gph } \partial f, \text{ while } f_k(x_k) \rightarrow f(x) \text{ for some sequence } x_k \rightarrow x \in \text{dom } f. \quad (2.22)$$

On the other hand,

$$\text{A sequence of closed proper convex functions } f_k \text{ epi-converges to such an } f \text{ if and only if their conjugate functions } f_k^* \text{ epi-converge to the conjugate } f^*. \quad (2.23)$$

In other words, the Legendre-Fenchel transform is continuous with respect to epi-convergence.

In general, it is possible for a sequence of functions to epi-converge without converging pointwise everywhere, and conversely. However, in the applications we will make involving random variables some degree of pointwise convergence can be utilized. This comes from the following characterization.

$$\text{For closed proper convex functions } f_k \text{ and } f \text{ on } (-\infty, \infty) \text{ with the same nonempty open interval } I \text{ as interior of } \text{dom } f \text{ and } \text{dom } f_k, \text{ the epi-convergence of } f_k \text{ to } f \text{ is equivalent to the pointwise convergence of } f_k \text{ to } f \text{ on the interval } I, \text{ or for that matter on a dense subset of } I, \text{ in which case the convergence is uniform on all compact subsets of } I. \quad (2.24)$$

Graphical convergence of maximal monotone relations can likewise be furnished with characterizations based on pointwise convergence:

$$\text{for maximal monotone } \Gamma_k \text{ and } \Gamma \text{ associated with nondecreasing functions } \gamma_k \text{ and } \gamma, \text{ graphical convergence of } \Gamma_k \text{ to } \Gamma \text{ corresponds to pointwise convergence of } \gamma_k \text{ to } \gamma \text{ at all continuity points of } \gamma, \text{ or equivalently, to pointwise convergence of } \gamma_k \text{ to } \gamma \text{ almost everywhere.} \quad (2.25)$$

On an open interval I where the functions γ_k and γ are finite, such pointwise convergence almost everywhere is furthermore equivalent to having

$$\int_a^b |\gamma_k(x) - \gamma(x)| dx \rightarrow 0 \quad \text{for all } [a, b] \subset I, \quad (2.26)$$

as seen through application of Lebesgue dominated convergence in the context of these functions being nondecreasing.

Second derivatives. Convex functions are known generally to be twice differentiable almost everywhere. Where does this enter our picture? Monotone relations provide a helpful graphical view.

For a closed proper convex function f the twice differentiability of f at $x \in \text{dom } f$ means that the one-sided derivative functions f'^- and f'^+ agree at x and are differentiable there. Graphically in terms of the monotone relation Γ giving ∂f , an equivalent statement is that there is a unique p such that $(x, p) \in \Gamma$, and furthermore a nonvertical tangent line to Γ at (x, p) .⁹ Here $f'(x) = p$ and the slope of the tangent is $f''(x)$. This slope must of course be nonnegative.

Let us call (x, p) a *nonsingular* point of Γ if there is a tangent line there which is nonvertical and also nonhorizontal. This corresponds to f having a *nonzero* second derivative at x . The symmetry in this notion provides us then with the following equivalences:

$$\begin{aligned} & f \text{ has } f'(x) = p \text{ and second derivative } f''(x) > 0 \\ \iff & (x, p) \text{ is a nonsingular point of } \Gamma \\ \iff & (p, x) \text{ is a nonsingular point of } \Delta = \Gamma^{-1}, \\ \iff & f^* \text{ has } f^{*'}(p) = x \text{ and second derivative } f^{*''}(p) > 0, \end{aligned} \quad (2.27)$$

in which case the second derivatives are reciprocal, $f^{*''}(p) = 1/f''(x)$.

Beyond passing to second derivatives in this manner, one can think of the maximal monotone relation Γ as directly associated with a measure “ $d\Gamma$ ” defined in Lebesgue-Stieltjes manner through the nondecreasing functions associated with it. (Vertical segments in Γ correspond to atoms in this measure, and the continuous nondecreasing function obtained by shrinking them out gives the rest of the measure in the usual way.) Likewise, the inverse relation Δ yields a measure “ $d\Delta$ ”. These measures are reciprocal in a certain sense that encompasses (2.27). They can be construed as the generalized second derivatives of f and f^* .

3 Back to Random Variables

We turn now to applying the general results in Section 2 to random variables in the setting of Section 1. We start with monotonicity and go on to duality. Then we see where this leads us with convergence issues. Supporting facts about expectations need to be recorded beforehand. To avoid complications that are inessential for our purposes, we make the assumption that

$$\text{henceforth all random variables } X \text{ have } E[|X|] < \infty. \quad (3.1)$$

⁹For tangency in general terms, see [26, Chapter 6].

Then $E[X]$ is well defined and finite, in particular.

Expectations with respect to the probability measure on $(-\infty, \infty)$ induced by a random variable X take the form of Lebesgue-Stieltjes integrals with respect to F_X . One has

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) dF_X(x) \quad (3.2)$$

for any (measurable) function g such that the integrand $g(x)$ is integrable with respect to the probability measure in question, or at least is bounded from below by something integrable; cf. Billingsley [3, Section 21].¹⁰ An expression of the same expectation in terms of the quantile function Q_X instead of the distribution function F_X is

$$E[g(X)] = \int_0^1 g(Q_X(p)) dp, \quad (3.3)$$

again as long as $g(Q_X(p))$ is bounded from below by something integrable. This holds because the integrals on the right of (3.2) and (3.3) agree through a change-of-variable rule; cf. Billingsley [3, Theorem 16.13].¹¹ In particular,

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x dF_X(x) = \int_0^1 Q_X(p) dp \quad (\text{finite}), \\ E[|X|^r] &= \int_{-\infty}^{\infty} |x|^r dF_X(x) = \int_0^1 |Q_X(p)|^r dp \quad \text{for } r \geq 1. \end{aligned} \quad (3.4)$$

Also conforming to this rule is the equivalence of the alternative definitions in (1.5) and (1.6) of the superquantile $\overline{Q}_X(p)$ for $p \in (0, 1)$. This equivalence can be identified with the version of the first equation in (3.4) that results from replacing F_X by the p -tail distribution function $F_X^{[p]}$ described right after (1.5) and replacing Q_X accordingly by the quantile function $Q_X^{[p]}$ for $F_X^{[p]}$, with $Q_X^{[p]}(t) = Q_X(p + t(1 - p))$ for $t \in (0, 1)$. Since $Q_X^{[p]}(t) \geq Q_X(p) > -\infty$, the integrand in the quantile integral is bounded from below by an integrable function on $(0, 1)$, and the equivalence between (1.5) and (1.6) is thereby justified.

Maximal monotonicity from distributions and quantiles. The distribution function F_X , which is nondecreasing right-continuous, has a left-continuous counterpart F_X^- . The monotonicity construction in Section 2, when applied to this pair, yields the relation Γ_X described in Section 1 in terms of “filling in the vertical gaps”:

$$\Gamma_X = \{ (x, p) \in \mathbb{R} \times \mathbb{R} \mid F_X^-(x) \leq p \leq F_X(x) \}. \quad (3.5)$$

Hence Γ_X is a *maximal monotone relation*. One can proceed similarly with the nondecreasing left-continuous function Q_X by extending it in the natural way beyond $(0, 1)$ with

$$Q_X(1) = \lim_{p \nearrow 1} Q_X(p), \quad Q_X(p) = \infty \text{ for } p > 1, \quad Q_X(p) = -\infty \text{ for } p \leq 0, \quad (3.6)$$

¹⁰Recall that the integral of a nonnegative (measurable) function is always well defined but might be ∞ . A function is “integrable” if its absolute value has finite integral. The integral of any function that is bounded below by an integrable function is likewise well defined as a finite value or ∞ .

¹¹The dF_X measure on $(-\infty, \infty)$ is the one induced from the dp measure on $(0, 1)$ by the function Q_X .

so as to get a nondecreasing left-continuous function on $(-\infty, \infty)$. Its extended right-continuous counterpart Q_X^+ has $Q_X^+(0) = \lim_{p \searrow 0} Q_X(p)$ and $Q_X^+(1) = \infty$. The relation Δ_X described in Section 1 is then

$$\Delta_X = \{ (p, x) \in \mathbb{R} \times \mathbb{R} \mid Q_X(p) \leq x \leq Q_X^+(p) \}, \quad (3.7)$$

and it, too, is therefore a *maximal monotone relation*. Moreover these relations are inverse to each other through the reciprocal formulas (1.2) and (1.3) for passing between F_X and Q_X :

$$(x, p) \in \Gamma_X \iff (p, x) \in \Delta_X, \quad \text{i.e., } \Delta_X = \Gamma_X^{-1} \text{ and } \Gamma_X = \Delta_X^{-1}. \quad (3.8)$$

This recalls the setting in (2.18) in which a pair of maximal monotone relations that are the inverses of each other are the graphs of the subdifferentials of two convex functions that are conjugate to each other. The construction of such functions are where we are now headed.

Superexpectation functions. A basic choice confronts us right away. We can pass from Γ_X to a convex function f having it as the graph of ∂f , but an additive constant is thereby left undetermined. An idea coming straight to mind is to look at $f(x) = \int_0^x F_X(x') dx'$, but that has a big disadvantage for applications, as will be explained below. Another choice with a lot behind it is taking f to be the function¹²

$$F_X^{(2)}(x) = E[\max\{0, x - X\}] = \int_{-\infty}^x F_X(x) dx, \quad (3.9)$$

which is finite¹³ and convex with right-derivative F_X . Ogryczak and Ruszcynski showed in [16, Theorem 3.1] that the conjugate of $F_X^{(2)}$ is the convex function given on $[0, 1]$ by¹⁴

$$F_X^{(-2)}(p) = \int_0^p Q_X(p') dp'. \quad (3.10)$$

but equalling ∞ outside of $[0, 1]$. It has Q_X as its left derivative on $(0, 1)$. In statistics, $F_X^{(2)}$ and $F_X^{(-2)}$ have long standing, but they emphasize the lower tail of X instead of the upper tail.

Desiring something tuned instead to upper tail properties, Dentcheva and Martinez in [4] introduced in parallel to (3.9) the “excess function”

$$H_X(x) = E[\max\{0, X - x\}] = \int_x^\infty [1 - F_X(x)] dx, \quad (3.11)$$

which likewise is finite and convex. They showed that its conjugate H_X^* can be expressed on $[0, 1]$ in terms of (although not directly as) the function¹⁵

$$L_X(p) = \int_p^1 Q_X(p') dp'. \quad (3.12)$$

However, L_X is concave, not convex, and it has $-Q_X(p)$ as its left-derivative at p , while H_X has $F_X(x) - 1$ as its right-derivative at x . Thus, this adaptation to a “cost” orientation of X does not sit comfortably in our duality framework.

¹²This function is traditionally important in “stochastic dominance,” to be taken up in Section 4.

¹³The finiteness of the integral is assured under (3.1).

¹⁴This gives the “Lorenz curve” [11] associated with X .

¹⁵This is the “upper Lorenz function” in their terminology, although the format of notation is ours.

A different choice will therefore be made here instead. It is dictated in part by our interest in coordinating with the “superquantiles” of random variables described in Section 1, as will be apparent when we come to duality.

We define the *superexpectation* function E_X associated with a random variable X by

$$E_X(x) = E[\max\{x, X\}] = \int_{-\infty}^{\infty} \max\{x, x'\} dF_X(x') = \int_0^1 \max\{x, Q_X(p)\} dp, \quad (3.13)$$

with the value $E_X(x)$ being termed the *superexpectation of X at level x* .¹⁶

Theorem 1 (characterization of superexpectations). *The superexpectation function E_X for a random variable X having $E[|X|] < \infty$ is a finite convex function on $(-\infty, \infty)$ which corresponds subdifferentially to the monotone relation Γ_X and the distribution function F_X through*

$$\Gamma_X = \text{gph } \partial E_X, \quad F_X(x) = E_X^{\uparrow}(x). \quad (3.14)$$

It is nondecreasing with

$$E_X(x) - x \geq 0, \quad \lim_{x \rightarrow \infty} [E_X(x) - x] = 0, \quad \lim_{x \rightarrow -\infty} E_X(x) = E[X] \quad (3.15)$$

and has the additional convexity property that

$$E_X(x) \leq (1 - \lambda)E_{X_0}(x) + \lambda E_{X_1}(x) \quad \text{when } X = (1 - \lambda)X_0 + \lambda X_1 \quad \text{with } 0 < \lambda < 1. \quad (3.16)$$

On the other hand, any convex function f on $(-\infty, \infty)$ with the properties that

$$f(x) - x \geq 0, \quad \lim_{x \nearrow \infty} [f(x) - x] = 0, \quad \lim_{x \searrow -\infty} f(x) = \text{a finite value}, \quad (3.17)$$

is E_X for a random variable X having $E[|X|] < \infty$.

Proof. The asymptotics in (3.15) are evident from $\max\{x, x'\} - x = \max\{0, x' - x\} \geq 0$, where the expressions as functions of x' decrease pointwise to 0 as x tends to ∞ but increase pointwise to x' as x tends to $-\infty$. To connect with F_X giving the right derivative, observe for $x' > x$ that

$$\frac{\max\{x', t\} - \max\{x, t\}}{x' - x} = \begin{cases} 1 & \text{if } t \leq x, \\ 0 & \text{if } t \geq x', \\ \frac{x' - t}{x' - x} \in (0, 1) & \text{if } t \in (x, x'), \end{cases}$$

and therefore

$$\text{prob}\{X \leq x\} \leq \frac{E_X(x') - E_X(x)}{x' - x} \leq \text{prob}\{X \leq x'\},$$

where the left side equals $F_X(x)$ and the right side equals $F_X(x')$. In taking the limit on both sides as $x' \searrow x$ and utilizing the right-continuity of F_X , we confirm that $E_X^{\uparrow}(x) = F_X(x)$.

The additional property in (3.16) is a consequence of the convexity of $\max\{x, X\}$ with respect to X in the definition (3.13) of $E_X(x)$.

If a convex function f has the properties in (3.17), it must be finite on $(-\infty, \infty)$ and nondecreasing. Moreover its left-derivatives $f'^-(x)$ and right-derivatives $f'^+(x)$ must lie in $[0, 1]$

¹⁶Relative to the excess function H_X in (3.11), we have clearly have $E_X(x) = H_X(x) + x$.

and increase to 1 as x tends to ∞ but decrease to 0 as x tends to $-\infty$. Thus in particular, the right-continuous function f'^+ meets the requirements of a distribution function F_X for a random variable X . \square

The properties in (3.17) say that the graph of E_X is above, but asymptotic to, the 45-degree line $y = x$. The additional convexity property in (3.16) is valuable for applications in stochastic optimization, which often involve random variables $X(u)$ that depend linearly or convexly on a parameter vector u . It is also a signal of the aptness of E_X as our designated choice of a convex function f having F_X as its right-derivative. This property also holds for $F_X^{(2)}$ as an antecedent of F_X , but that choice concentrates on the lower tail instead of the upper tail. It is absent for other seemingly natural choices, such as $f(x) = \int_a^x F_X(x')dx'$. In that case,

$$\int_a^x F_X(x')dx' = E_X(x) - E_X(a) \quad \text{for any } a \in (-\infty, \infty),$$

since both sides the same right-derivatives in x and both vanish at a . Although $E_X(a)$, like $E_X(x)$, is convex with respect to X , the difference $E_X(x) - E_X(a)$ that property.

Conjugate superexpectations. Dualization of the superexpectation function E_X through the Legendre-Fenchel transform will be addressed next. This is where the superquantiles $\bar{Q}_X(p)$ of (1.5) and (1.6) come on stage.

The conjugacy claim in the following theorem is new only in its formulation, in view of the conjugacy between (3.9) and (3.10) already established by Ogrychak and Ruszczyński in [16], and the result of Dentcheva and Martinez in [4] about the relationship between the functions in (3.11) and (3.12). However, the proof we supply takes a different route.

Theorem 2 (dualization of superexpectations). *The closed proper convex function E_X^* on $(-\infty, \infty)$ that is conjugate to the superexpectation function E_X on $(-\infty, \infty)$ is given by*

$$E_X^*(p) = \begin{cases} -(1-p)\bar{Q}_X(p) & \text{for } p \in (0, 1), \\ -E[X] & \text{for } p = 0, \\ 0 & \text{for } p = 1, \\ \infty & \text{for } p \notin [0, 1]. \end{cases} \quad (3.18)$$

It is continuous relative to $[0, 1]$, entailing

$$\lim_{p \nearrow 1} (1-p)\bar{Q}_X(p) = 0, \quad \lim_{p \searrow 0} \bar{Q}_X(p) = E[X], \quad (3.19)$$

and it corresponds subdifferentially to the maximal monotone relation $\Delta_X = \Gamma_X^{-1}$ and the quantile function Q_X through

$$\Delta_X = \text{gph } \partial E_X^*, \quad Q_X(p) = E_X^{*'}(p). \quad (3.20)$$

On the other hand, any function g on $(-\infty, \infty)$ that is finite convex and continuous on $[0, 1]$ with $g(1) = 0$, but $g(p) = \infty$ for $p \notin [0, 1]$, is E_X^* for some random variable X .

Proof. Let g denote the function of $p \in (-\infty, \infty)$ described by the right side of (3.18). It will be demonstrated in steps that g is a closed proper convex function having E_X as its conjugate

g^* . That will tell us through (2.16) that g is in turn the conjugate E_X^* . It will also confirm the limits in (3.19), since a closed proper convex function on $(-\infty, \infty)$ is always continuous relative to an interval on which it is finite, cf. [17, Corollary 7.5.1].

From the expression for $\overline{Q}_X(p)$ in (1.6), already justified as being equivalent to the one in (1.5), we have $g(p) = -\int_p^1 Q_X(p')dp'$ for $p \in (0, 1)$. This implies that $g^-(p) = Q_X(p)$ and $g^+(p) = Q_X^+(p)$ on $(0, 1)$. Since the limit of $-\int_p^1 Q_X(p')dp'$ as $p \nearrow 1$ is 0, while the limit as $p \searrow 0$ is $-\int_0^1 Q_X(p')dp' = -E[X]$ by (3.4), g is continuous relative to $[0, 1]$. Since Q_X is nondecreasing, g is also convex on $[0, 1]$ and hence, in its extension outside of $[0, 1]$ by ∞ , is a closed proper convex function on $(-\infty, \infty)$. Furthermore, the left- and right-derivative functions for g , as extended in the manner of (2.14), are the functions Q_X and Q_X^+ as extended in (3.6). The graph of ∂g , as determined by definition from g^- and g^+ , is therefore the relation Δ_X in (3.7).

It follows then from (2.18) that the convex function g^* conjugate to g has the relation $\Gamma = \Delta^{-1}$ as the graph of ∂g^* . Since E_X is already known from Theorem 1 to have Γ as the graph of ∂E_X , the functions g^* and E_X can differ at most by a constant, $E_X = g^* + c$. On taking conjugates again, we get $E_X^* = (g^* + c)^* = (g^*)^* - c = g - c$. Thus, to verify that $c = 0$, confirming that $E_X^* = g$, it will suffice to show that $E_X^*(1) = 0$. For this we apply the formula for the Legendre-Fenchel transform: $E_X^*(p) = \sup_x \{ px - E_X(x) \}$ at $p = 1$. This gives us

$$-E_X^*(1) = \inf_x \{ -x + E[\max\{x, X\}] \} = \inf_x \{ E[\max\{0, X - x\}] \},$$

where the expectation of $\max\{0, X - x\}$ is always ≥ 0 but approaches 0 as $x \rightarrow \infty$.

For the last part of the theorem, we note that for any g as described there the function $q = g^-$ on $(0, 1)$ is left-continuous and nondecreasing, with $g(p) = -\int_p^1 q(p')dp'$. In other words, q meets the requirement of being a quantile function Q_X for which the right side of (3.18) can be identified with g . Then g must be the corresponding function E_X^* . \square

Corollary (superquantile functions). *The conjugate E_X^* is uniquely determined by the superquantile function \overline{Q}_X . Not only it, but also E_X , F_X and Q_X , along with Γ_X and Δ_X , can be reconstructed from knowledge of \overline{Q}_X . Moreover the following properties of a function \bar{g} on $(0, 1)$ are necessary and sufficient to have $\bar{g} = \overline{Q}_X$ for a random variable X with $E[|X|] < \infty$:*

$$(1 - p)\bar{g}(p) \text{ is concave in } p \text{ with } \lim_{p \nearrow 1} (1 - p)\bar{g}(p) = 0, \quad \lim_{p \searrow 0} \bar{g}(p) = \text{a finite value.} \quad (3.21)$$

Proof. Once \overline{Q}_X has determined E_X^* from (3.18), we get E_X as the conjugate $(E_X^*)^*$. These functions yield F_X and Q_X through one-sided differentiation, and we then have the monotone relations Γ_X and Δ_X as well. The conditions listed for a function \bar{g} correspond to the conditions on g at the end of Theorem 2. \square

Besides this characterization, it is interesting to observe as a consequence of the formula (1.6) for superquantiles $\overline{Q}_X(x)$ that

$$\begin{aligned} \overline{Q}_X \text{ is a continuous increasing function of } p \in (0, 1) \text{ with} \\ \overline{Q}_X^-(p) = \frac{\overline{Q}_X(p) - Q_X(p)}{1 - p} \leq \frac{\overline{Q}_X(p) - Q_X^+(p)}{1 - p} = \overline{Q}_X^+(p). \end{aligned} \quad (3.22)$$

In contrast, Q_X is only nondecreasing, not (strictly) increasing, and can be discontinuous. There is no assurance that Q_X has left-derivatives or right-derivatives apart from the general dictum that a nondecreasing function is differentiable almost everywhere.

Example (exponential distributions). Let X be exponentially distributed with parameter $\lambda > 0$. Then the distribution function is $F_X(x) = 1 - \exp(-\lambda x)$, the superexpectation function is

$$E_X(x) = \begin{cases} x + (1/\lambda) \exp(-\lambda x) & \text{for } x \geq 0, \\ 1/\lambda & \text{for } x < 0, \end{cases}$$

and the conjugate superexpectation function has $E_X^*(p) = (1/\lambda)(p - 1)(1 - \log(1 - p))$ for $p \in [0, 1)$. Quantiles and superquantiles are thus given on $(0, 1)$ by

$$Q_X(p) = -(1/\lambda) \log(1 - p), \quad \bar{Q}_X(p) = (1/\lambda)(1 - \log(1 - p)).$$

Our results further make available new estimates for work with superquantiles.

Theorem 3 (superquantile estimates). For $p \in [0, 1)$, one has

- (a) $|\bar{Q}_X(p) - \bar{Q}_Y(p)| \leq \frac{1}{1-p} E|X - Y|$ when $E[|X|] < \infty$, $E[|Y|] < \infty$.
- (b) $E[X] \leq \bar{Q}_X(p) \leq E[X] + \frac{1}{\sqrt{1-p}} \sigma(X)$ when $E[X^2] < \infty$, $\sigma(X) = \text{standard deviation}$.

Proof. Observe first that $|\max\{x, a\} - \max\{x, b\}| \leq |a - b|$. For the superexpectation functions corresponding to X and Y , this gives us

$$|E_X(x) - E_Y(x)| = E[|\max\{x, X\} - \max\{x, Y\}|] \leq E[|X - Y|].$$

or in other words, both $E_Y \leq E_X + E[|X - Y|]$ and $E_X \leq E_Y + E[|X - Y|]$. Applying the Legendre-Fenchel transform, which reverses functional inequalities, we see that

$$E_Y^* \geq E_X^* - E[|X - Y|], \quad E_X^* \geq E_Y^* - E[|X - Y|],$$

and consequently $|E_X^*(p) - E_Y^*(p)| \leq E[|X - Y|]$ for $p \in [0, 1)$. Then (3.18) yields (a).

For (b), we note that $-E_X^*(p) = \int_p^1 Q_X(p') dp' = \int_0^1 Q_X(p') I_{[p,1]}(p') dp'$ for the characteristic function of the interval $[1, p]$, while recalling that $E[X] = \inf \bar{Q}_X$. Then, by way of (3.18), we have $0 \leq (1 - p)(\bar{Q}_X(p) - E[X]) = \int_0^1 (Q_X(p') - E[X]) I_{[p,1]}(p') dp'$. The Cauchy-Schwartz inequality provides now that

$$(1 - p)(\bar{Q}_X(p) - E[X]) \leq \left[\int_0^1 (Q_X(p')(p') - E[X])^2 dp' \right]^{1/2} \left[\int_0^1 I_{[p,1]}(p')^2 dp' \right]^{1/2},$$

where the first factor on the right is $(E[|X - E[X]|^2])^{1/2} = \sigma(X)$ by (3.3) and the second is $\sqrt{1 - p}$. In dividing through by $1 - p$, one gets the upper bound in (b). \square

Convergence in distribution. Convergence of a sequence of random variables X_k to a random variable X can now be brought into focus. There are several concepts of importance, but the one we concentrate on is *convergence in distribution*, which is customarily defined by¹⁷

$$X_k \rightarrow X \text{ in distribution when } F_{X_k}(x) \rightarrow F_X(x) \text{ at all continuity points } x \text{ of } F_X. \quad (3.23)$$

¹⁷In probability theory, random variables are generally presented as functions on a probability space. The convergence of a sequence is viewed then with all the random variables regarded as functions on the same probability space. Here we are working directly with distributions and only nominally with particular random variables giving rise to them, which are not unique. However, Skorohod's theorem reconciles these points of view; cf. [3, Theorem 25.6]. It says that when distribution functions F_k converge to a distribution function F in the manner of (3.23), it is possible to construct random variables X_k and X on a common probability space such that $F_k = F_{X_k}$, $F = F_X$, and the functions X_k converge pointwise to X .

This classical property has various characterizations, for instance

$$X_k \rightarrow X \text{ in distribution} \iff E[g(X_k)] \rightarrow E[g(X)] \text{ for bounded continuous } g, \quad (3.24)$$

which is recorded in Billingsley [3, Theorem 25.8]. Here we provide characterizations beyond such classical theory.

Theorem 4 (characterizations of convergence in distribution). *For a sequence of random variables X_k , convergence in distribution to a random variable X is equivalent also to each of the following conditions:*

- (a) Γ_{X_k} converges graphically to Γ_X ,
- (b) Δ_{X_k} converges graphically to Δ_X ,
- (c) $Q_{X_k}(p) \rightarrow Q_X(p)$ at all continuity points p of Q_X in $(0, 1)$,
- (d) $E_{X_k}(x) \rightarrow E_X(x)$ for all $x \in (-\infty, \infty)$,
- (e) $\bar{Q}_{X_k}(p) \rightarrow \bar{Q}_X(p)$ for all $p \in (0, 1)$.

Proof. The equivalence with (a) is evident from the description of graphical convergence in (2.25). The equivalence with (b) then follows because graphical convergence is preserved when taking inverses. Application of (2.25) to the convergence in (b) gives the equivalence with (c).

When the defining property in (3.23) holds, the integrals $\int_x^\infty [1 - F_{X_k}(x')] dx'$ converge to $\int_x^\infty [1 - F_X(x')] dx'$ (inasmuch as the integrands are uniformly bounded). This yields the property in (d) through the fact that $E_X(x) = H_X(x) + x$ for the function H_X in (3.11). For the opposite implication, if (d) holds we can use derivative estimates for convex functions in the form

$$\frac{E_X(x) - E_X(x - \epsilon)}{\epsilon} \leq E_X^{\prime-}(x) \leq E_X^{\prime+}(x) \leq \frac{E_X(x + \epsilon) - E_X(x)}{\epsilon} \quad (3.25)$$

for any $\epsilon > 0$ and in parallel

$$\frac{E_{X_k}(x) - E_{X_k}(x - \epsilon)}{\epsilon} \leq E_{X_k}^{\prime-}(x) \leq E_{X_k}^{\prime+}(x) \leq \frac{E_{X_k}(x + \epsilon) - E_{X_k}(x)}{\epsilon}. \quad (3.26)$$

where $E_{X_k}^{\prime+}(x) = F_{X_k}(x)$ and $E_X^{\prime+}(x) = F_X(x)$. At a continuity point x of F_X we also have $E_X^{\prime-}(x) = F_X(x)$. Since the outer bounds in (3.25) approach those in (3.26) as $k \rightarrow \infty$ by (d), we conclude that

$$\frac{E_X(x) - E_X(x - \epsilon)}{\epsilon} \leq \liminf_{k \rightarrow \infty} F_{X_k}(x) \leq \limsup_{k \rightarrow \infty} F_{X_k}(x) \leq \frac{E_X(x + \epsilon) - E_X(x)}{\epsilon} \quad (3.27)$$

The upper and lower bounds in (3.27) both converge to $E_X^{\prime+}(x) = F_X(x)$ at the continuity point x , and therefore $F_{X_k}(x) \rightarrow F_X(x)$. Thus, (d) is equivalent to the defining property in (3.23) for convergence in distribution.

Next we observe that, since (d) concerns finite convex functions, the pointwise convergence there is equivalent to the epi-convergence of E_{X_k} to E_X ; recall (2.24). Applying (2.23) we get the epi-convergence of the conjugate functions $E_{X_k}^*$ to E_X^* , and then the equivalence with (e), once more via (2.24). \square

By taking advantage of (2.24), the everywhere pointwise convergence in (d) and (e) can be replaced by pointwise convergence on a dense subset or uniform convergence on compact

intervals of $(-\infty, \infty)$ and $(0, 1)$, respectively. On the other hand, the pointwise convergence property in (c) can be elaborated in terms of the alternative descriptions in (2.25) and (2.26).

It is apparent from (e) that a superquantile is stable under perturbations of the underlying probability distribution. This has importance consequences for optimization problems with superquantiles of parametric random variables as objective functions and constraints. If the superquantiles remain convex and finite as functions of the parameters, then Theorem 4 with (2.24) ensures epiconvergence of approximations obtained by replacing true probability distributions with approximating ones. Moreover, optimal solutions of problems with approximations will tend to those of the true problems, justifying the use of approximate probability distributions in applications.

Other useful implications of convergence in distribution, which relax the boundedness of g in (3.24), can be derived from conditions on moments. Let us say, for $r \geq 1$, that a continuous function $g : (-\infty, \infty) \rightarrow (-\infty, \infty)$ has *growth rate at most r* when $\lim_{|x| \rightarrow \infty} |g(x)|/|x|^r < \infty$. This is equivalent to having $c > 0$ such that $|g(x)| \leq c(1 + |x|^r)$ everywhere.

Theorem 5 (further properties of convergence in distribution). *If X_k converges in distribution to X and $\limsup_k E[|X_k|^{r(1+\epsilon)}] < \infty$ for some $r \geq 1$ and $\epsilon > 0$, then*

$$E[g(X_k)] \rightarrow E[g(X)] \text{ (finite) for continuous } g \text{ having growth rate at most } r. \quad (3.28)$$

Proof. Consider $Y_k(p) = g(Q_{X_k}(p))$ and $Y(p) = g(Q_X(p))$ as random variables on the probability space $(0, 1)$. We have $E[Y_k] = E[g(X_k)]$, $E[Y] = E[g(X)]$, by (3.3) and know from Theorem 4 that the convergence in distribution of X_k to X entails Y_k as a function on $(0, 1)$ converging pointwise to Y almost everywhere. Our aim is to show that the growth assumptions imply $E[Y_k] \rightarrow E[Y]$ with $E[Y]$ finite. For that it suffices to confirm that those assumptions guarantee uniform integrability of the functions Y_k in the sense that

$$\lim_{a \rightarrow \infty} \sup_k \int_{|Y_k| \geq a} |Y_k(p)| dp = 0, \quad (3.29)$$

see Billingsley [3, Theorem 25.12]. Because g has growth rate $\leq r$, there exists $c > 0$ such that $|Y_k(p)| \leq c(1 + |Q_{X_k}(p)|^r)$ for all k . It will be enough therefore to confirm that

$$\lim_{b \rightarrow \infty} \sup_k \int_{Z_k \geq b} Z_k(p) dp = 0 \text{ for } Z_k(p) = |Q_{X_k}(p)|^r. \quad (3.30)$$

We have (via Billingsley [3, (25.13)]) the estimate for any $\epsilon > 0$ that

$$\int_{Z_k \geq b} Z_k(p) dp \leq \frac{1}{b^\epsilon} \int_0^1 Z_k^{1+\epsilon}(p) dp = \frac{1}{b^\epsilon} \int_0^1 |Q_{X_k}(p)|^{r(1+\epsilon)} dp = \frac{1}{b^\epsilon} E[|X_k|^{r(1+\epsilon)}].$$

Under our assumption that the expectations $E[|X_k|^{r+\epsilon}]$ are uniformly bounded from above (for k sufficiently large), we obtain (3.30) and the desired uniform integrability. \square

As a particular case of Theorem 5 one can take $g(x) = |x|^r$ in (3.28) to get convergence of moments: $E[|X_k|^r] \rightarrow E[|X|^r]$. Note that even $E[X_k] \rightarrow E[X]$ is not assured by convergence in distribution of X_k to X without something extra, despite having $\overline{Q}_{X_k}(p) \rightarrow \overline{Q}_X(p)$ almost everywhere with $E[X_k] = \inf \overline{Q}_{X_k}$ and $E[X] = \inf \overline{Q}_X$. Here the sufficient condition given for $E[X_k] \rightarrow E[X]$ is the boundedness of $E[|X_k|^{1+\epsilon}]$ as $k \rightarrow \infty$ for some $\epsilon > 0$.

Distribution densities and quantile densities. The symmetric view of second derivatives of convex functions and their conjugates, built at the end of Section 3 around the maximal monotone relations associated with them, will now be applied to random variables.

If a distribution function F_X is differentiable, its derivative F'_X gives the *distribution density function* f_X for X . Then¹⁸

$$\int_{-\infty}^{\infty} g(x) dF_X(x) = \int_{-\infty}^{\infty} g(x) f_X(x) dx. \quad (3.31)$$

What is new now is the perspective from Theorem 1 that $f_X(x)$ is the second derivative $E''_X(x)$, and that a sort of duality lies in the background.

The measure $d\Gamma_X = dF_X$ has a counterpart $d\Delta_X = dQ_X$, the Lebesgue-Stieltjes measure associated with the quantile function Q_X as a nondecreasing left-continuous function on $(0, 1)$. We can equally contemplate the differentiability of Q_X , interpreting it as yielding a *quantile density function* q_X on $(0, 1)$, with $q_X(p)$ being the second derivative $E_X^{*''}(p)$ according to Theorem 2. Then¹⁹

$$\int_0^1 h(p) dQ_X(p) = \int_0^1 h(p) q_X(p) dp. \quad (3.32)$$

It is interesting in this respect to note that, through change of variables,²⁰ one has

$$\int_0^1 h(p) dQ_X(p) = \int_{-\infty}^{\infty} h(F_X(x)) dx. \quad (3.33)$$

This is the quantile version of the distribution rule in the equivalence between (3.2) and (3.3).

Full differentiability of F_X and Q_X is not a prerequisite to all insights. The available facts can be specialized without that, although full differentiability does produce the nicest picture.

Theorem 6 (duality of densities). *The following relations hold in general:*

$$\begin{aligned} &F_X \text{ has derivative } F'_X(x) > 0 \\ &\iff (x, p) \text{ is a nonsingular point of } \Gamma_X \\ &\iff (p, x) \text{ is a nonsingular point of } \Delta_X = \Gamma_X^{-1}, \\ &\iff Q_X \text{ has derivative } Q'_X(p) > 0 \\ &\text{in which case the derivatives are reciprocal, } Q'_X(p) = 1/F'_X(x). \end{aligned} \quad (3.34)$$

In consequence,

$$\begin{aligned} &F_X \text{ is differentiable on } (-\infty, \infty) \text{ with } F'_X(x) > 0 \text{ for } x \in (\inf X, \sup X), \\ &\iff Q_X \text{ is differentiable on } (0, 1) \text{ with } Q'_X(p) > 0 \text{ for } p \in (0, 1), \end{aligned} \quad (3.35)$$

in which case

$$\begin{aligned} &Q'_X(p) = 1/F'_X(Q_X(p)) \text{ for } p \in (0, 1), \\ &F'_X(x) = 1/Q'_X(F_X(x)) \text{ for } x \in (\inf X, \sup X). \end{aligned} \quad (3.36)$$

Proof. All of this is immediate from (2.27) with F'_X and Q'_X being the second derivatives of the convex functions E_X and E_X^* . \square

¹⁸For measurable functions g that are integrable with respect to the dF_X measure.

¹⁹For measurable functions h that are integrable with respect to the dQ_X measure.

²⁰Again applying the rule in Billingsley [3, Theorem 16.13]; the dQ_X measure on $(0, 1)$ is the one induced from the dx measure on $(-\infty, \infty)$ by the function F_X .

4 Applications to Quantifying Risk

The importance of Q_X and \bar{Q}_X as so-called measures of risk has been recalled in Section 1, but more can be said with the facts now at our disposal. An explanation of the joint minimization formula (1.11) for $Q_X(p)$ and $\bar{Q}_X(p)$ will be taken up first. An extension to a parallel formula, in which $\bar{Q}_X(p)$ gives the argmin instead of the min, will follow.

Derivation of the joint rule for quantiles and superquantiles. The proof of Theorem 2 took shortcuts by utilizing facts in convex analysis, but a direct approach to calculating E_X^* from E_X by the Legendre-Fenchel transform was the original route to discovery of the minimization formula (1.11). The motivation in the first place was to obtain a minimization formula for quantiles based on knowing that Γ_X is the graph of ∂E_X , namely

$$[Q_X(p), Q_X^+(p)] = \partial E_X^{-1}(p) = \partial E_X^*(p) = \operatorname{argmin}_x \{ E_X(x) - xp \} \text{ for any } p \in (0, 1). \quad (4.1)$$

by (2.17) and (2.18). Here

$$E_X(x) - px = (1-p)x + (E[\max\{x, X\}] - x) = (1-p)\left(x + \frac{1}{1-p}E[\max\{0, X-x\}]\right),$$

and consequently $[Q_X(p), Q_X^+(p)]$ is the set of x 's that minimize $x + \frac{1}{1-p}E[\max\{0, X-x\}]$. The argmin part of (1.11) is just this. At the same time we see that the Legendre-Fenchel formula $E_X^*(p) = \sup_x \{ px - E_X(x) \}$, with attainment guaranteed for $p \in (0, 1)$, translates to

$$-\frac{E_X^*(p)}{1-p} = \min_x \left\{ x + \frac{1}{1-p}E[\max\{0, X-x\}] \right\} \text{ for } p \in (0, 1).$$

The left side is $\bar{Q}_X(p)$ by Theorem 2, and this clinches the other half of the rule in (1.11).

Extension to “higher-order superquantiles.” We proceed now to look for an analog of (1.11) in which the superquantiles take the place of quantiles in giving the minimum. The reason for wanting to do this is the role of quantiles, and potentially superquantiles, in generalized regression of the kind considered in [24] and [25], but explaining all that here would carry us far away from the current theme. “Superquantile regression” is the subject introduced and developed in our paper [21], with support from results secured here.

An observation to start from is that the main term $E[\max\{0, X\}]$ in (1.11) has the additional expressions

$$E[\max\{0, X\}] = \int_{-\infty}^{\infty} \max\{0, x\} dF_X(x) = \int_0^1 \max\{0, Q_X(p)\} dp. \quad (4.2)$$

It turns out that all we need to do in order to build the right analog of (1.11) is to replace F_X by a different but closely related distribution function \bar{F}_X such that

$$\text{the quantiles of } \bar{F}_X \text{ are the superquantiles of } F_X. \quad (4.3)$$

As indicated graphically in Section 1, this *superdistribution* function is obtained by “inverting” \bar{Q}_X , namely

$$\bar{F}_X(x) = \begin{cases} \bar{Q}_X^{-1}(x) & \text{for } \lim_{p \searrow 0} \bar{Q}_X(p) < x < \lim_{p \nearrow 1} \bar{Q}_X(p), \\ 0 & \text{for } x \leq \lim_{p \searrow 0} \bar{Q}_X(p), \\ 1 & \text{for } x \geq \lim_{p \nearrow 1} \bar{Q}_X(p). \end{cases} \quad (4.4)$$

It is the distribution function for the random variable \bar{X} associated with X by (1.7), so that

$$Q_{\bar{X}}(p) = \bar{Q}_X(p).$$

Much that has already been worked out for F_X carries over to \bar{F}_X , as long as $E[|\bar{X}|] < \infty$ in accordance with the blanket assumption in (3.1) that we have been relying on. That is the case when $E[X^2] < \infty$, as seen through the estimate in Theorem 3(b). In particular, then,

$$\bar{F}_X = \bar{E}'_X \quad \text{for} \quad \bar{E}_X(x) = \int_{-\infty}^{\infty} \max\{x, x'\} d\bar{F}_X(x') = \int_0^1 \max\{x, \bar{Q}_X(p)\} dp, \quad (4.5)$$

where the equivalence holds as an echo of (3.12) in the face of (4.3). The function \bar{E}_X is finite and convex on $(-\infty, \infty)$, again with $\bar{E}_X(x) - x$ positive and tending to 0 as $x \rightarrow \infty$.

The conjugate function \bar{E}_X^* can be determined by applying Theorem 2 in this setting. The main ingredient in the resulting formula is the replacement of $\bar{Q}_X(p)$ by a higher analog, namely

$$\bar{\bar{Q}}_X(p) = \frac{1}{1-p} \int_{\bar{Q}_X(p)}^{\infty} x' d\bar{F}_X(x') = \frac{1}{1-p} \int_p^1 \bar{Q}_X(p') dp'. \quad (4.6)$$

This “supersuperquantile” is the conditional expectation of \bar{X} in its p -tail with respect to the $\bar{F}_X = F_{\bar{X}}$ distribution. The complications with the original definition of the p -tail fall away because \bar{F}_X has no jumps; the \bar{F}_X distribution has no “probability atoms.” With respect to \bar{F}_X , the interval $[\bar{Q}_X(p), \infty)$ has probability $1-p$. As a matter of fact, $\bar{\bar{Q}}_X = \bar{Q}_{\bar{X}}$.

This suggests, through (4.2), that the analog of the expression \mathcal{V}_p in (1.11) as a “measure of regret” might be taken to be

$$\bar{\mathcal{V}}_p(X) = \frac{1}{1-p} \int_{-\infty}^{\infty} \max\{0, x\} d\bar{F}_X(x) = \frac{1}{1-p} \int_0^1 \max\{0, \bar{Q}_X(p')\} dp', \quad (4.7)$$

and this does indeed give us what we want.

Theorem 7 (superquantiles as quantiles). *Suppose $E[|X|^2] < \infty$. Then, as a measure of risk, $\mathcal{R}(X) = \bar{\bar{Q}}_X(p)$ has the coherency properties in (1.10), like $\mathcal{R}(X) = \bar{Q}_X(p)$. In terms of $\bar{\mathcal{V}}_p$ defined in (4.7), the two can be calculated simultaneously for $p \in (0, 1)$ by*

$$\begin{aligned} \bar{\bar{Q}}_X(p) &= \operatorname{argmin}_x \{x + \bar{\mathcal{V}}_p(X - x)\}, \\ \bar{Q}_X(p) &= \min_x \{x + \bar{\mathcal{V}}_p(X - x)\}. \end{aligned} \quad (4.8)$$

The functional $\bar{\mathcal{V}}_p$ itself has the regret properties of \mathcal{V}_p in (1.12), specifically

$$\begin{aligned} \bar{\mathcal{V}}_p(X) &\leq \bar{\mathcal{V}}_p(X') \quad \text{when } X \leq X' \text{ almost surely,} \\ \bar{\mathcal{V}}_p(X + X') &\leq \bar{\mathcal{V}}_p(X) + \bar{\mathcal{V}}_p(X'), \\ \bar{\mathcal{V}}_p(\lambda X) &= \lambda \bar{\mathcal{V}}_p(X) \quad \text{for } \lambda > 0, \\ \bar{\mathcal{V}}_p(X) &\geq E[X], \quad \text{with equality holding only when } X \equiv 0. \end{aligned} \quad (4.9)$$

Proof. The parallel structure suffices to confirm (4.8). The coherency properties of $\mathcal{R}(X) = \bar{\bar{Q}}_X(p)$ in (1.10) with respect to X lead through the second integral expression in (4.6) to

those same properties holding for $\mathcal{R}(X) = \overline{Q}_X(p)$. The properties in (4.9) similarly come from invoking (1.10) for $\overline{Q}_X(p)$ in the second formula for $\overline{V}_p(X)$ in (4.7) and calling on the fact that $\overline{Q}_X(p)$ is an increasing function with $E[X]$ as its infimum. \square

The minimization in (4.8) may seem to demand too much knowledge of the regret functional \overline{V}_p be practical, but properties of the superquantile integrand, such as the estimates in Theorem 3, can come to the rescue. The elementary theory of integration (approximation of integrands by step functions or piecewise linear functions) leads to approximating expressions for $\overline{V}_p(X)$ that come from linear combinations of superquantiles $\overline{Q}_X(p_k)$. The formula for $\overline{Q}_X(p)$ in (1.11) for any p can be employed to calculate the value of such an expression for any X . Upper and lower estimates can be developed for the closeness of such an expression to $\mathcal{V}_p(X)$. Such estimates are worked out in our paper [21].

Stochastic dominance. Another notion that enters the study of risk is stochastic dominance. Two versions, known as first-order and second-order, are especially important, but the issue of “usage orientation” of a random variable again has to be respected. Most often, stochastic dominance is articulated for the context of a random variable X being preferable to a random variable Y when its outcomes are, by some quantification standard, generally higher. That is profit/gain orientation, but in this article we are treating cost/loss/damage orientation, so some inequalities need to be reversed in identifying the “dominance” of X over Y with X being “better” than Y .

In profit/gain/benefit orientation, it is customary to define first-order stochastic dominance $X \geq_1 Y$ as corresponding to $F_X \leq F_Y$ (the graph of F_X therefore being to the right of the graph of F_Y). Second-order stochastic dominance $X \geq_2 Y$ is taken as $F_X^{(2)} \leq F_Y^{(2)}$; cf. (3.10). It is well known that these properties translate into having $E[g(X)] \geq E[g(Y)]$ for a class of increasing functions in the first case and a class of increasing concave functions in the second. However, some authors prefer to take such expectation properties directly as the definition, since they provide the main motivation for the concept in applications. We follow that pattern here in adapting to cost/loss/damage orientation.

Definition (first- and second-order dominance, inverted). *First-order stochastic dominance of X over Y in cost/loss/damage orientation, to be denoted by $X \leq'_1 Y$ here,²¹ and second-order stochastic dominance, $X \leq'_2 Y$, mean the following:*

$$\begin{aligned} X \leq'_1 Y &\iff E[g(X)] \leq E[g(Y)] \text{ for continuous bounded increasing } g, \\ X \leq'_2 Y &\iff E[g(X)] \leq E[g(Y)] \text{ for finite convex increasing } g. \end{aligned} \tag{4.12}$$

Recall here that a finite convex function g is automatically continuous. Also, it always has $g(x) \geq ax + b$ for some $a > 0$ and $b \in (-\infty, \infty)$, so that the expectations in (4.12) are sure to be well defined, although possibly ∞ , but not $-\infty$ (under our blanket assumption (3.1) on finite expectations).

If g is interpreted as a penalty function, the inequalities in (4.12) concern expected penalties under X and Y . The two conditions then describe situations involving a pair of cost/loss random variables X and Y in which X is less risky than Y regardless of the particular penalty

²¹The prime is a reminder of the switch from the usual orientation as seen in textbooks.

function g that may have to be faced—within some category.²² This is attractive in situations where a decision maker may have little knowledge of the penalties. An important example for stochastic dominance in the profit/gain orientation comes up in finance, where penalty functions are replaced by utility functions and convexity in the second-order case by concavity.

The second property in (4.12) is also known as “increasing convex order,” \leq_{ic} , cf. [12], and was featured by Dentcheva and Martinez [4] in their adaptation to cost/loss orientation.

Theorem 8 (stochastic dominance in cost/loss/damage orientation). *First-order stochastic dominance is characterized by*

$$X \leq'_1 Y \iff F_X \geq F_Y \iff Q_X \leq Q_Y. \quad (4.13)$$

Second-order stochastic dominance is characterized by

$$X \leq'_2 Y \iff E_X \leq E_Y \iff \bar{Q}_X \leq \bar{Q}_Y. \quad (4.14)$$

Proof. We rely here, in part, on characterizations in the profit/gain orientation furnished by Föllmer and Schied [7] (and elsewhere). Their Theorem 2.70 covers (4.13) with a slight difference coming from our focus on the left-continuous quantile function. They contemplate a class of “quantile functions” between these and their right-continuous partners, and accordingly replace the pointwise inequality $Q_X \leq Q_Y$ by an almost everywhere inequality.

For (4.14) the derivation is a bit more complicated and specialized toward the concepts in this article. The cost/loss version of one characterization of second-order dominance in Theorem 2.70 of [7] is that

$$X \leq'_2 Y \iff E[\max\{0, X - c\}] \leq E[\max\{0, Y - c\}] \text{ for all } c. \quad (4.15)$$

Because $E_X(c) = E[\max\{c, X\}] = c + E[\max\{0, X - c\}]$ and similarly $E_Y(c)$, we can translate (4.15) to saying that $E_X(c) \leq E_Y(c)$ for all c . The observation to make next is that the Legendre-Fenchel transform converts $E_X \leq E_Y$ to $E_X^* \geq E_Y^*$. The formula in Theorem 2 lets us identify this with $\bar{Q}_X \leq \bar{Q}_Y$. \square

Stochastic dominance has important applications to constraint modeling in stochastic optimization; see Dentcheva and Ruszczyński [5].

Comonotonicity. Another way that monotone relations enter the framework of risk is through the property of comonotonicity.

Definition (comonotonicity of random variables). *Two random variables X_1 and X_2 are said to be comonotone if the essential range of outcomes of the pair (X_1, X_2) is a monotone relation Γ in $\mathbb{R} \times \mathbb{R}$.*²³

This means roughly that the two random variables move in tandem; the risk in one cannot hedge against the risk in the other. Indeed, it implies the existence of a third random variable X along with increasing Lipschitz continuous functions f_1 and f_2 such that $X_1 = f_1(X)$ and

²²This explains why we use \leq in (4.12). It signals “less risky.”

²³The essential range is the smallest closed set that, with probability 1, contains all outcomes.

$X_2 = f_2(X)$. For this, one can simply take $X = X_1 + X_2$ and apply the Minty parameterization of a maximal extension of Γ ; cf. (2.3).

Besides the motivation for comonotonicity as capturing this tandem behavior of a pair of random variables, there are consequences for their quantiles and superquantiles. The fact that comonotonicity of random variables leads to additivity of their quantiles, the initial property below, is well known; cf. [7, Lemma 4.84]. We offer an argument for the converse and indicate how this ties in with superquantiles and superexpectations.

Theorem 9 (characterizations of comonotonicity). *The following properties of a pair of random variables X_1 and X_2 are equivalent to comonotonicity:*

- (a) $Q_{X_1+X_2}(p) = Q_{X_1}(p) + Q_{X_2}(p)$ for all $p \in (0, 1)$,
- (b) $\overline{Q}_{X_1+X_2}(p) = \overline{Q}_{X_1}(p) + \overline{Q}_{X_2}(p)$ for all $p \in (0, 1)$,
- (c) $E_{X_1+X_2}(x) = \min_{x_1+x_2=x} \{E_{X_1}(x_1) + E_{X_2}(x_2)\}$ for all $x \in (-\infty, \infty)$.

Proof. First we suppose comonotonicity and show that then (a) holds. The monotonicity of the essential range Γ of (X_1, X_2) makes the function $\phi : (x_1, x_2) \rightarrow x_1 + x_2 = x$ map Γ monotonically one-to-one into the real line. The joint probability distribution of X_1 and X_2 on $\mathbb{R} \times \mathbb{R}$, concentrated in Γ , is thereby transformed into the probability distribution of $X = X_1 + X_2$, concentrated in $\phi(\Gamma)$. For any $p \in (0, 1)$ the quantile $Q_X(p)$ gives the highest point x of $\phi(\Gamma)$ such that $F_X(x) \leq p$. The unique antecedent $\phi^{-1}(x) = (x_1(x), x_2(x)) \in \Gamma$ then has to be $(Q_{X_1}(p), Q_{X_2}(p))$. Thus, $Q_X(p) = Q_{X_1}(p) + Q_{X_2}(p)$, as claimed.

To demonstrate the converse, that (a) implies comonotonicity, we can make use of the fact that the essential range of a random variable X is the closure of the range of its quantile function Q_X . It is traced by $Q_X(p)$ as p goes from 0 to 1 in $(0, 1)$, except that where jumps occur the right limit $Q_X^-(p)$ needs also to be brought in. This can be invoked for X_1 , X_2 and $X = X_1 + X_2$ to see that, when (a) holds, the probability parameter p traces the range of (X_1, X_2) monotonically as $(Q_{X_1}(p), Q_{X_2}(p))$. This range is then a monotone relation.

The equivalence between (a) and (b) is obvious from the formula (1.6) for superquantiles in terms of quantiles. This yields a further equivalence through Theorem 2 with having

$$E_{X_1+X_2}^*(p) = E_{X_1}^*(p) + E_{X_2}^*(p) \text{ for all } p.$$

Applying the rule in convex analysis that the conjugate of a sum is obtained by the operation $\#$ of ‘‘inf-convolution’’ on the conjugate functions,²⁴

$$(E_{X_1}^* + E_{X_2}^*)^*(x) = (E_{X_1}^{**} \# E_{X_2}^{**})(x) = \inf_{x_1+x_2=x} \{E_{X_1}^{**}(x_1) + E_{X_2}^{**}(x_2)\},$$

we arrive at (c). □

Theorem 9 relates also to an associated concept of comonotonicity for measures of risk due to Ogryczak and Ruszczyński [14], [15], [16], namely that

$$\mathcal{R} \text{ is comonotonic if } \mathcal{R}(X_1 + X_2) = \mathcal{R}(X_1) + \mathcal{R}(X_2) \text{ when } X_1, X_2, \text{ are comonotone.} \tag{4.16}$$

²⁴For more on inf-convolution, see [17] and [26].

The theorem says, among other things, that

$$\text{the risk measure } \mathcal{R}(X) = \overline{Q}_X(p) \text{ is comonotonic for every } p \in (0, 1). \quad (4.17)$$

It is easy to see that this carries over also to the mixed superquantile measures of risk in (4.10) and (4.11). More on this topic can be found in the book of Föllmer and Schied [7] in coordination with applications in finance.

Acknowledgement: This work was supported by the Air Force Office of Scientific Research under Grants FA9550-11-1-0206 and F1ATAO1194GOO1.

References

- [1] ACERBI, C., AND TASCHE, D. “On the coherence of expected shortfall,” *Journal of Banking and Finance* **26** (2002), 1487–1503.
- [2] ARTZNER, P., DELBAEN, F., EBER, J.-M., AND HEATH, D., “Coherent measures of risk,” *Mathematical Finance* **9** (1999), 203–227.
- [3] BILLINGSLEY, P., *Probability and Measure*, Wiley, anniversary edition: 2012.
- [4] DENTCHEVA, D. AND MARTINEZ, G., “Two-stage stochastic optimization problems with stochastic ordering constraints on the recourse,” *European Journal of Operational Research* **219** (2012), 1–8.
- [5] DENTCHEVA, D. AND RUSZCZYŃSKI, A., “Stochastic optimization with dominance constraints,” *SIAM Journal on Optimization* **14** (2003), 548–566.
- [6] DENTCHEVA, D. AND RUSZCZYŃSKI, A., “Common mathematical foundations of expected utility and dual utility theories,” *SIAM Journal on Optimization* **24** (2013).
- [7] FÖLLMER, H. AND SCHIED, A., *Stochastic Finance*, 2nd edition, De Gruyter, New York, 2004.
- [8] KOENKER, R. AND BASSETT, G., “Regression quantiles,” *Econometrica* **46** (1978), 33–50.
- [9] KOENKER, R., *Quantile Regression*, Econometric Society Monograph Series, Cambridge University Press, 2005.
- [10] KUSUOKA, S., “On law invariant coherent risk measures,” *Advances in Mathematical Economics* **3** (2001), 83–95.
- [11] LORENZ, M., “Methods of measuring concentration of wealth,” *Journal of the American Statistical Association* **9** (1905), 209–219.

- [12] MILLER, A. AND STOYAN, D., *Comparison Methods for Stochastic Models and Risks*, Wiley, 2002.
- [13] MINTY, G. J., “Monotone networks,” *Proceedings Royal Soc. London A* **257** (1960), 194–212.
- [14] OGRYCZAK, W. AND RUSZCZYŃSKI, A., “From stochastic dominance to mean-risk models: semideviations and risk measures,” *European Journal of Operations Research* **116** (1999), 33–50.
- [15] OGRYCZAK, W. AND RUSZCZYŃSKI, A., “On consistency of stochastic dominance and mean-semideviation models,” *Mathematical Programming* **89** (2001), 217–232.
- [16] OGRYCZAK, W. AND RUSZCZYŃSKI, A., “Dual stochastic dominance and related mean-risk models,” *SIAM Journal on Optimization* **13** (2002), 60–78.
- [17] ROCKAFELLAR, R. T., *Convex Analysis*, Princeton Univ. Press, Princeton, NJ., 1970.
- [18] ROCKAFELLAR, R. T., *Conjugate Duality and Optimization*, No. 16 in the Conference Board of Math. Sciences Series, SIAM, Philadelphia, 1974.
- [19] ROCKAFELLAR, R. T., *Network Flows and Monotropic Optimization*, Wiley, 1984; re-published by Athena Scientific, 1998.
- [20] ROCKAFELLAR, R. T., AND ROYSET, J. O., “On buffered failure probability in design and optimization of structures,” *Journal of Reliability Engineering and System Safety* **99** (2010), 499–510.
- [21] ROCKAFELLAR, R. T., ROYSET, J. O., AND MIRANDA, S. I., “Superquantile Regression with Applications to Buffered Reliability, Uncertainty Quantification, and Conditional Value-at-Risk,” *Working paper, Naval Postgraduate School, Monterey, CA* (2012).
- [22] ROCKAFELLAR, R. T. AND URYASEV, S., “Optimization of conditional value-at-risk,” *Journal of Risk* **2** (2000), 21–42.
- [23] ROCKAFELLAR, R. T. AND URYASEV, S., “Conditional value-at-risk for general loss distributions,” *Journal of Banking and Finance* **26** (2002), 1443–1471.
- [24] ROCKAFELLAR, R. T. AND URYASEV, S. “The fundamental risk quadrangle in risk management, optimization and statistical estimation,” *Surveys in Operations Research and Management Science* (2013).
- [25] ROCKAFELLAR, R. T., URYASEV, S., AND ZABARANKIN, M., “Risk tuning with generalized linear regression,” *Math. of Operations Research* **33** (2008), 712–729.
- [26] ROCKAFELLAR, R. T., AND WETS, R. J-B, *Variational Analysis*, Springer-Verlag, New York, 1998