



HAL
open science

Risk Quantization by Magnitude and Propensity

Olivier P. Faugeras, Gilles Pages

► **To cite this version:**

Olivier P. Faugeras, Gilles Pages. Risk Quantization by Magnitude and Propensity. 2021. hal-03233068v2

HAL Id: hal-03233068

<https://hal.science/hal-03233068v2>

Preprint submitted on 25 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Risk Quantization by Magnitude and Propensity

Olivier P. Faugeras^{*1} and Gilles Pagès²

¹Toulouse School of Economics, Université Toulouse 1 Capitole, 1, Esplanade de l'Université. Bureau T106. 31080 Toulouse Cedex 06, France.

`olivier.faugeras@tse-fr.eu`

²Laboratoire de Probabilités, Statistique et Modélisation, UMR 8001, Campus Pierre et Marie Curie, Sorbonne Université, case 158, 4, pl. Jussieu, F-75252 Paris Cedex 5, France. `gilles.pages@sorbonne-universite.fr`

August 21, 2022

Abstract

We propose a novel approach in the assessment of a random risk variable X by introducing magnitude-propensity risk measures (m_X, p_X) . This bivariate measure intends to account for the dual aspect of risk, where the magnitudes x of X tell how high are the losses incurred, whereas the probabilities $P(X = x)$ reveal how often one has to expect to suffer such losses. The basic idea is to simultaneously quantify both the severity m_X and the propensity p_X of the real-valued risk X . This is to be contrasted with traditional univariate risk measures, like VaR or CVaR, which typically conflate both effects. In its simplest form, (m_X, p_X) is obtained by mass transportation in Wasserstein metric of the law of X to a two-points $\{0, m_X\}$ discrete distribution with mass p_X at m_X . The approach can also be formulated as a constrained optimal quantization problem. This allows for an informative comparison of risks on both the magnitude and propensity scales. Several examples illustrate the usefulness of the proposed approach. Some variants, extensions and applications are also considered.

Keywords: magnitude-propensity; risk measure; mass transportation; optimal quantization.

1 Introduction and outline

The evaluation and comparison of risks are basic tasks of risk analysis. The usual view in Insurance mathematics is to evaluate an univariate risk¹ X by a risk measure

^{*}Corresponding author

¹We take the Insurance Mathematics convention, where the non-negative values of X stands for the loss incurred

$\rho(X)$, which can be thought of as a one-sided deterministic univariate summary of the random variable X . Risks X, Y are then compared through their respective risk measures $\rho(X), \rho(Y)$.

The starting point of this paper is the basic realization that risk, as a random variable, is intrinsically a bivariate phenomenon: magnitudes (loss amounts) occurs with given propensities (or probabilities). Hence, it appears inescapable that a risk measure, as a single univariate quantity on the magnitude scale, will conflate both effects, thus giving a somehow blurred representation of the risk borne by the random variable X . It would then be of interest to quantify risk on both the magnitude and propensity scales. The purpose of this paper is to propose such a simultaneous quantification.

The proposed approach is based on the following idea: as mentioned above, a risk measure $\rho(X)$ can be viewed as a deterministic proxy of the random risk X . Distributionally speaking, it can be thought of as a Dirac measure $\delta_{\rho(X)}(\cdot)$ at $\rho(X)$: this distribution gives full propensity one at the magnitude level $\rho(X)$. Therefore, in order to quantify the risk with a varying propensity p_X , it makes sense to look for an approximate of the distribution of X by a mixture of two Dirac, a Dirac at location zero with weight $1 - p_X$, and a Dirac at location m_X , with weight p_X . This proxy distribution thus encodes the magnitude and propensity effect of the risk borne by X , through the pair (p_X, m_X) . Mathematically, this problem of approximating distributions is carried out by mass transportation in Wasserstein metric. Optimal transportation to discrete measure also corresponds to the problem of optimal quantization, well-known in the Engineering and Signal Processing literature. Hence, the proposed approach to quantify risk on both the magnitude and propensity scales amounts to a special, constrained, optimal quantization problem.

The outline of the paper is as follows: In Section 2, we motivate the magnitude - propensity approach to risk measures. As a new paradigm to risk evaluation, the proposed approach needs a careful and detailed exposition of its main idea. We first argue and give some evidence of this magnitude-propensity duality of risk, using (intentionally) simplistic examples to make our point clear. We then show how some classical risk measures typically mixes both effects and explain why it would be desirable to quantify risk on both the magnitude and propensity scales. A risk measure can usually be derived as a M-functional, i.e. as a statistical parameter which is a solution of a problem of minimization of some expected loss. We next show how these expected loss minimization problems can be embedded into (degenerate) optimal transportation problems, i.e. mass transportation towards a Dirac distribution. Eventually, we propose our definition of the magnitude-propensity pair (p_X, m_X) , as hinted above, and make the connection to optimal quantization.

Section 3 is a theoretical study of the basic idea. We show how to compute the optimal (m_X, p_X) , both as direct minimization problem and as an optimal quantization problem. To that purpose we briefly recall the main facts about quantization theory. We provide existence, characterization and (partial) uniqueness results. We discuss some basic properties of the obtained magnitude-propensity pairs and study some example distributions.

In Section 4, we give some numerical illustrations of the basic idea. We explain how the optimal (m_X, p_X) gives rise to magnitude-propensity plots, which allow for an informative comparison of risks on both the magnitude and propensity scale. We show how to do such comparisons for distributions with or without explicit formulas for (m_X, p_X) . Empirically, the optimal magnitude-propensity pair can be computed by a fixed point algorithm, which is akin to Lloyd's algorithm in optimal quantization.

The methodology is illustrated on a real data set of insurance losses.

Eventually, Section 5 broadens the perspective on the proposed approach by sketching some variants, extensions and possible applications to Insurance. In particular, we extend the risk quantization to several points to account for the “tail risk” of a distribution, in addition to its “moderate risk”. We next discuss several possible variants to accommodate for real-valued risks, thus giving a magnitude-propensity quantization of both the losses and the gains of a distribution. Then, multivariate generalizations are studied empirically on a set of distributions and show the influence of the dependence structure and the marginals on the multivariate magnitude-propensity measures. Next, we give an application to portfolio optimization and show how the proposed magnitude-propensity measures can be used to define acceptable risk regions and select the best or worst portfolio. At last, we mention possible extensions using covariates. We conclude in Section 6.

2 The magnitude-propensity approach to risk measures

2.1 Risk is intrinsically a bivariate propensity-magnitude random phenomenon

As a primary concept of Insurance Mathematics, it is somehow difficult to give a precise, deductive definition of risk in terms of more primitive concepts. This explains why, e.g. Novak (2012) p. 224, states that “There is currently no consensus concerning the meaning of the word risk”. Hence, risk is often given an implicit or semantic definition in textbooks, as in Novak (2012) p. 191: “Risk is a possibility of an undesirable event. Though such an event is rare, its magnitude can be devastating.”

The latter description of risk encompasses the duality magnitude-propensity of random variables: intuitively, a claim can be “risky” because losses may happen often, i.e. with a “high” propensity, possibly with (relatively) “small” magnitudes, or because a catastrophe of very large magnitude may happen, albeit with a (relatively) “low” propensity. We mention that this dual nature of risk is explicit in the Engineering literature (see e.g. Bedford and Cooke (2001)), where it is summarized by the semantic formula “risk=uncertainty+damage” in Kaplan and Garrick (1981).

This duality of risk manifests itself on several levels:

1. Intrinsic magnitude-propensity aspect for a single random risk

When one considers risk as a property of a single r.v., i.e. as being intrinsic, a claim X can be deemed risky due to high magnitudes, or due to high propensities, as explained above. It is easy to construct examples of two risky claims X_1 , X_2 , with same average risk $\mathbb{E} X_1 = \mathbb{E} X_2$, yet with X_1 having high magnitude risk, while X_2 having high propensity risk.

2. Relative magnitude-propensity effect for a pair of risks

One can also consider that risk is not intrinsic but a relative property of one r.v. X_1 w.r.t. another r.v. X_2 , hence a property of the pair (X_1, X_2) , see Faugeras and Rüschendorf (2018) for the notion of risk excess measures. The magnitude-propensity duality then manifests itself in the comparison of risks: a claim X_2 can be riskier than X_1 , but is it so due to a difference in magnitudes, or due to a difference in propensities (or both)? Yet, the risk excess measures of Faugeras and Rüschendorf (2018) typically conflates both effects.

3. Magnitude-propensity effect in the stochastic order

In turn, the magnitude-propensity duality encountered in comparing risks translates in the ordering structure of probability measures. We recall that the usual stochastic order \prec_{st} (a.k.a. first-order stochastic dominance) is defined as,

$$X_1 \prec_{st} X_2 \iff F_{X_1}(x) \geq F_{X_2}(x), \quad \forall x \in \mathbb{R}, \quad (1)$$

$$\iff Q_{X_1}(t) \leq Q_{X_2}(t), \quad \forall t \in (0, 1), \quad (2)$$

where F_X , resp. $Q_X = F_X^{-1}$, stands for the c.d.f., resp. quantile function (generalized inverse), of X . It can be described either as a comparison of magnitudes, as expressed via the quantile functions in (2), or as a comparison of propensities, as expressed via the c.d.f.s in (1).

In the following toy example (intentionally simplistic to illustrate the point), we show that the stochastic order can conflate both types of differences:

Toy example 1. (a) Let X_1 s.t. $P(X_1 = 0) = 0.9$, $P(X_1 = 100) = 0.1$, while X_2 is s.t. $P(X_2 = 0) = 0.9$, $P(X_2 = 1000) = 0.1$. Then, X_1 and X_2 have same propensities, but X_2 is more risky than X_1 , solely due to a difference in magnitudes.

(b) Let X_1 s.t. $P(X_1 = 0) = 0.95$, $P(X_1 = 1000) = 0.05$, while X_2 is s.t. $P(X_2 = 0) = 0.5$, $P(X_2 = 1000) = 0.5$. Then, X_1 and X_2 have same magnitudes, but X_2 is more risky than X_1 , solely due to a difference in propensities.

In both cases, (a) and (b), one has $X_1 \prec_{st} X_2$, as is shown on their c.d.f.s in Figure 1. Here, the stochastic comparison does not distinguish between a horizontal shift of c.d.f.s due to a difference in magnitudes (Figure 1 left), and a vertical shift of c.d.f.s due to a difference of propensities (Figure 1 right).

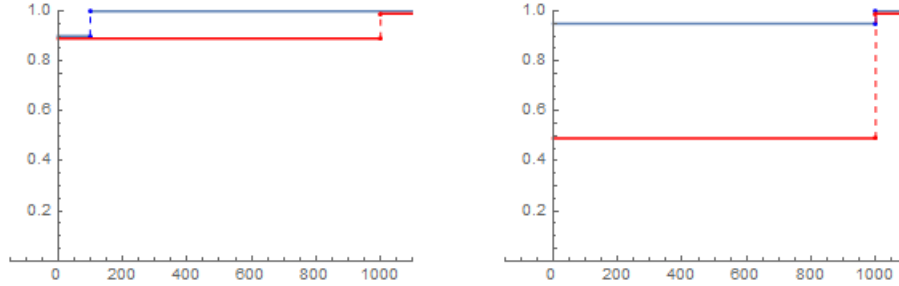


Figure 1: C.d.f.s of X_1 (blue), X_2 (red). Left: case (a), difference of magnitudes (horizontal shift). Right: case (b), differences of propensities (vertical shift). In both cases, $X_1 \prec_{st} X_2$.

2.2 Risk measures as univariate deterministic proxies for a random risk

The main paradigm to the assessment of risk in Insurance and Financial Mathematics is based on the use of risk measures

$$\rho: \mathcal{X} \rightarrow \mathbb{R}^+, \quad (3)$$

where \mathcal{X} is a space of non-negative measurable random variables modeling an insurance claim. These risk measures quantify risk on a magnitude scale, by converting a random loss X into a deterministic certainty equivalent $\rho(X)$. The latter can then be used for ordering different risks, and for decision making purposes, like setting the premium for covering the risk X , see e.g. Rüschenendorf (2013). The resulting risk measure should typically satisfy some desirable properties (coherent risk measure), like translation invariance, monotonicity, etc., see e.g. Artzner et al. (1999), Föllmer and Schied (2002).

The last decades have seen a multiplication of risk measures. Among the numerous approaches encountered in the literature, let us mention the classical premium calculation principles based on probabilistic models (see e.g. Mikosch (2009), Asmussen and Albrecher (2010), Bühlmann (1996)), the axiomatic approach where premium principles are subject to a set of desirable properties (see Artzner et al. (1999)), the abstract/functional analytic approach where risk measures are derived from an acceptance set and a set of scenario measures (see Föllmer and Schied (2002)), distortion-based measures Wang (1996), and eventually the approach to risk excess measures induced by hemi-metrics (see Fageras and Rüschenendorf (2018)). These numerous approaches have led to considerable debate on the pros and cons of the risk measures available in the literature, see e.g. Embrechts et al. (2014).

From the magnitude-propensity point of view, the duality of risk is reflected in its measurement, i.e. in the risk measures themselves. Some risk measures focus more on the propensity aspect of risk, while some others focus more on the magnitude one, and most risk measures mix both. Let us illustrate this point with the following comparison of three well-known risk measures:

- The Value-At-Risk, defined as the left α -quantile of X ,

$$VaR_\alpha(X) := \inf\{x : P(X \leq x) \geq \alpha\}, \quad 0 < \alpha < 1 \quad (4)$$

encodes a propensity into a magnitude, by setting the $VaR_\alpha(X)$ as the utmost-left α -quantile. (Note that some authors define VaR as the utmost-right quantile). Value-at-Risk only controls the probability of a loss, it does not capture the size of such a loss if it occurs.

- The opposite extreme is, for X an essentially bounded random variable, the essential supremum,

$$\rho_\infty(X) := \text{ess sup } X,$$

which quantifies the maximum magnitude of the loss, but gives no information on the probabilities.

- The Conditional Value-at-Risk, see e.g. Uryasev and Rockafellar (2001) or Kaas et al., 2008, is defined as

$$CVaR_\alpha(X) := \frac{1}{1-\alpha} \int_\alpha^1 VaR_\alpha(t) dt \quad (5)$$

which reduces for a continuous distribution to the Conditional Tail Expectation (CTE),

$$CVaR_\alpha(X) = CTE_\alpha(X) := \mathbb{E}[X | X > VaR_\alpha(X)], \quad (6)$$

see e.g. Definition 2.15 and Lemma 2.16 in McNeil, Frey, and Embrechts (2005)².

²Note that in the latter reference CVaR is called Expected Shortfall, and that other terms abound in the literature such as Tail Value-at-Risk (TVaR), Average Value-at-Risk (AVaR), Expected Tail Loss (ETL), Superquantile, etc.

CVaR thus clearly mixes the two aspects magnitude-propensity of the risk by computing a weighted average over a threshold.

In general, one wants to know both when/how often a catastrophe may occur, and also what is the size/extent of the loss one has to face. This suggests that reducing risk X to a single proxy on the magnitude scale as a univariate risk measure $\rho(X)$ is somehow inadequate to account for the dual nature of risk. This idea that one numerical quantity cannot hedge against risk has already been evoked beforehand in the literature, see e.g. Rootzén and Klüppelberg (1999). Let us also remark that in their criticism of the VaR measure, the authors of the academic response to the Basel 3.5 framework are implicitly interested in these dual propensity and magnitude aspects of risk (see Embrechts et al. (2014) p.27 “Question W1: VaR says nothing concerning the what-if question: Given we encounter a high loss, what can be said about its magnitude?”). It thus would be of considerable practical interest to quantify risk on both the magnitude and propensity scales. This is the purpose of this paper.

Remark 1 (On elicibility). *It has been argued in the literature that elicibility is also a desirable property for risk measures (see Embrechts et al. (2014)). Elicibility is a concept derived from point forecasting. Roughly speaking, in order that a point forecast, derived from a statistical functional, be consistent with their evaluation by averaging over the past, the statistical functional must be written as a (strict) M -functional, see Gneiting (2011) for a precise definition. It is known that VaR is not a coherent risk measure, but is elicitable, while the Conditional Value-at-Risk (CVaR) is a coherent law invariant risk measure, but is not elicitable. CVaR is jointly elicitable with VaR, see Fissler and Ziegel (2016).*

This consideration of joint elicibility and coherence of (CVaR, VaR), i.e. of combining magnitude and propensity type univariate risk measures, is another supplementary motivation for arguing that one should switch to bivariate risk measures for the study of univariate risks.

Remark 2 (On parametrized risk measures). *For risk measures depending on a parameter, like α in $VaR_\alpha(X)$ or $CVaR_\alpha(X)$, arise the practical issue of choosing the “right” parameter value α for correctly representing the risk by a single numerical quantity. Common practice is to hedge against a “rare event”, i.e. to take, say, $\alpha = 0.9, 0.95$ or 0.99 .*

VaR and CVaR give in fact curves, $\alpha \rightarrow VaR_\alpha(X)$ and $\alpha \rightarrow CVaR_\alpha(X)$. These curves, under mild conditions, determine the distribution of X . Hence, they give the same information as the c.d.f. F_X . They are just another possible analytical characterization of the distribution of X . A complete assessment of the magnitude and propensity effects of a risk X can be visualised by plotting its c.d.f. F_X , or more conveniently its survival function \bar{F}_X , possibly on a log-log scale. This is the classical approach in the Engineering literature, see Kaplan and Garrick (1981), where it is argued that “a single number is not a big enough concept to communicate risk. It takes a whole curve”.

However, it becomes difficult to compare entire curves. Therefore, it is natural to look for a summary of this distribution-determining curve to as few as possible numerical quantities. This is the path favoured in Insurance and Financial Mathematics with risk measures. The magnitude-propensity measure to be introduced below, can thus be viewed as a middle-ground between univariate risk measures of Insurance Mathematics and the full curve approach in Engineering.

2.3 M-statistical functionals can be obtained from mass transportation to a degenerate distribution

The following discussion gives the key insight for defining a magnitude-propensity risk measure as an optimal transportation problem. Let us recall that the Monge-Kantorovich optimal transportation problem aims at finding a joint measure $P^{(X,Y)}$ on, say, the product measurable space $(\mathbb{R} \times \mathbb{R}, \mathcal{B}(\mathbb{R}^2))$, with prescribed marginals (P^X, P^Y) , which is the solution of the optimisation problem:

$$\mathcal{T}_c(P^X, P^Y) := \inf_{P^{(X,Y)} \in \mathcal{P}(P^X, P^Y)} \int c(x, y) P^{(X,Y)}(dx, dy), \quad (7)$$

where $c : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^+$ is a cost function and the infimum is on the set $\mathcal{P}(P^X, P^Y)$ of joint distributions $P^{(X,Y)}$ with given marginals P^X, P^Y . Informally, mass at x of P^X is transported to y , according to the conditional distribution $P(dy|x)$ of the optimal transportation plan $P^{(X,Y)}$, in order to recover P^Y while minimising the average cost of transportation $\int c(x, y) P^{(X,Y)}(dx, dy)$. Under regularity conditions, the optimal transportation plan is induced by a (Monge) mapping T , viz. $P^{(X,Y)} = P^{(X, T(X))}$. See Rachev and Rüschendorf (1998), Villani (2003), Villani (2009), Santambrogio (2015) for book-length treatment on the subject.

When P^Y is degenerate, i.e. $P^Y = \delta_m$ with $m \in \mathbb{R}$, then $\mathcal{P}(P^X, \delta_m)$ reduces to the singleton product measure $\mathcal{P}(P^X, \delta_m) = \{P^X(dx) \times \delta_m(dy)\}$. (7) thus simplifies as the expected cost between $X \sim P^X$ and a fixed point m ,

$$\mathcal{T}_c(P^X, \delta_m) = \int c(x, m) P^X(dx) = \mathbb{E} c(X, m). \quad (8)$$

Therefore, minimizing the transportation cost (7) over the set $\mathcal{D} := \{\delta_m(dy), m \in \mathbb{R}\}$ of Dirac measures is equivalent to minimizing the expected cost (8) over $m \in \mathbb{R}$,

$$\mathcal{T}_c(P^X, \mathcal{D}) := \inf_{P^Y \in \mathcal{D}} \mathcal{T}_c(P^X, P^Y) = \inf_{m \in \mathbb{R}} \mathbb{E} c(X, m). \quad (9)$$

In particular,

- for $c(x, y) = (x - y)^2$, (7) is the squared L_2 -Wasserstein metric W_2 and (9) writes as the variance,

$$W_2^2(P^X, \mathcal{D}) = \inf_{P^Y \in \mathcal{D}} W_2^2(P^X, P^Y) = \inf_{m \in \mathbb{R}} \mathbb{E}(X - m)^2 = \text{Var}(X),$$

and is obviously minimised for the mean $m = EX$. In addition, when P^X is replaced by the conditional law of X given $X > \text{VaR}_\alpha(X)$, one obtains the CVaR (5).

- For the L_1 distance, $c(x, y) = |x - y|$, one gets the median.
- For the asymmetric cost $c(x, y) = (x - y)\alpha \mathbb{1}_{x-y \geq 0} + (y - x)(1 - \alpha)\mathbb{1}_{y-x > 0}$, with $0 < \alpha < 1$, one obtains the (left)- α -quantile, $m = q_\alpha(X)$ (see e.g. Koenker (2005)), that is to say the Value-At-Risk (4).
- For $c(x, y) = y + \frac{(x-y)\mathbb{1}_{x \geq y}}{1-\alpha}$, one gets simultaneously the Conditional Value-at-Risk and the Value-at-Risk, when $\mathbb{E}[X] < \infty$: the optimal value in (9) is the Conditional Value-at-Risk while the argmin gives the Value-At-Risk, see Uryasev and Rockafellar (2001), Rockafellar and Uryasev (2002).

- And so on for other statistical functionals. In particular, see Faugeras and Rüschendorf (2018) for risk excess measures obtained by optimal transportation with cost functions which are hemi-metrics encoding an order.

Summarizing the above discussion, statistical functionals and risk measures, which can be expressed as an M-estimator solving (9) for a suitable cost function, can be regarded as being obtained from a special mass transportation problem towards a family of degenerate Dirac distribution δ_m .

2.4 The magnitude-propensity (m_X, p_X) approach to measuring risk

The optimal transportation view on risk measures of Section 2.3 suggests that the limitations of the risk measures of Section 2.2 come from the fact that these univariate functionals can be viewed as being obtained by mass transportation from the P^X measure to a degenerate Dirac δ_m measure: the latter distribution only bears a magnitude m with full propensity. Hence, the magnitude and propensity aspects of P^X are mixed and encoded in the sole magnitude m of the Dirac destination measure.

It therefore becomes natural to suggest a mass transportation approach to risk measures, as in (9), but with the target Dirac distribution δ_m replaced by a two-points distribution P^Y ,

$$P^Y = (1 - p)\delta_0 + p\delta_m. \quad (10)$$

The latter distribution encodes both the magnitude and propensity aspects of risk: a loss of magnitude m occurs with probability p (and no loss occurs with probability $1 - p$). We can therefore define of the basic idea of the paper as follows:

Definition 2.1. *Let $\mathcal{A}_0 := \{P^Y = (1 - p)\delta_0 + p\delta_m, \quad p \in (0, 1), m \in \mathbb{R}^+\}$ be the set of such two-point distributions (10). For $X \sim P^X$ with $\mathbb{E}[X^2] < \infty$, the bivariate magnitude-propensity risk measure (m_X, p_X) is obtained by minimizing the Wasserstein W_2 distance from P^X to \mathcal{A}_0 ,*

$$(m_X, p_X) = \arg \inf_{P^Y \in \mathcal{A}_0} W_2(P^X, P^Y). \quad (11)$$

In other words, the risk X is approximated in Wasserstein metric by a proxy $Y \sim B_0$, which bears a loss of magnitude m_X , occurring with a probability p_X , and no loss with probability $1 - p_X$. We exclude the values $p = 0$, $p = 1$ and $m = 0$ in the family \mathcal{A}_0 , in order to obtain non degenerate measures.

This approach of approximating a distribution by a discrete one corresponds to the well-known problem of optimal quantization, see Graf and Luschgy (2000). The latter is itself related to k -means clustering, see Pollard (1982). Here, the main difference is that one location of the discrete distribution is constrained to be 0. This point mass at zero encodes the absence of loss and thus the point mass p_X at $m_X > 0$ encodes loss. $P^Y \in \mathcal{A}_0$ is characterized by the two parameters (m, p) , and quantization to \mathcal{A}_0 gives the minimal way to represent the magnitude and propensity effect borne by X .

3 Theoretical analysis

The theoretical study of the optimal (m_X, p_X) magnitude-propensity pair in (11) can be done by two main approaches: either as an optimal mass transportation problem or as an optimal quantization problem. The more refined results are obtained by

the latter approach. However, we first make use of the former to quickly obtain a characterisation of the optimal magnitude-propensity pair.

3.1 Computation of (m_X, p_X) by the optimal transportation approach

A direct optimization approach can be used to characterize the (m_X, p_X) by using the explicit form of the Wasserstein metric in dimension one: denote by Q_X the quantile function of P^X ,

$$Q_X(t) := \inf\{x : F_X(x) \geq t\}, \quad 0 < t < 1.$$

It is then well-known (see e.g. Rachev and Rüschendorf (1998)) that

$$W_2^2(P^X, P^Y) = \int_0^1 (Q_X(t) - Q_Y(t))^2 dt, \quad (12)$$

for univariate P^X, P^Y with finite variance. For $Y \in \mathcal{A}_0$, its quantile function writes $Q_Y(t) = m\mathbb{1}_{1-p < t \leq 1}$. Plugging the later in (12) and optimizing gives the following characterization result.

3.1.1 Characterization by direct optimisation

In the remainder of the paper, we simplify notations and will denote simply by F , Q , f the c.d.f., quantile function and density (if it exists) of X .

Proposition 3.1. *Let X s.t. $\mathbb{E}[X^2] < \infty$.*

i) Necessary conditions: if (m_X, p_X) is a local minimum of (11), then it satisfies the following system of equations,

$$m_X = 2Q(1 - p_X), \quad (13)$$

$$m_X = \frac{\int_{1-p_X}^1 Q(t) dt}{p_X}. \quad (14)$$

ii) Sufficiency condition: if P^X is absolutely continuous with continuous density f , then (m_X, p_X) is optimal if $f(Q(1 - p_X)) > 0$ and

$$\frac{2p_X}{f(Q(1 - p_X))} - Q(1 - p_X) > 0. \quad (15)$$

Proof. i) The squared Wasserstein distance between P^X and $P^Y \in \mathcal{A}_0$ writes

$$\begin{aligned} W_2^2(P^X, P^Y) &= \int_0^{1-p} (Q(t))^2 dt + \int_{1-p}^1 (Q(t) - m)^2 dt \\ &= E[X^2] + m^2 p - 2m \int_{1-p}^1 Q(t) dt \\ &:= \psi(m, p) \end{aligned}$$

ψ is differentiable and any optimal magnitude-propensity (m_X, p_X) solving (11) must satisfy the first order conditions

$$\begin{cases} \frac{\partial \psi(m_X, p_X)}{\partial m} = 0 \\ \frac{\partial \psi(m_X, p_X)}{\partial p} = 0 \end{cases} \Leftrightarrow \begin{cases} -2 \int_{1-p_X}^1 Q(t) + 2m_X p_X = 0 \\ -2m_X Q(1 - p_X) + m_X^2 = 0. \end{cases}$$

For $m_X \neq 0$ and $p_X \neq 0$, one gets (13) and (14).

- ii) if P^X has density f such that $f(Q(p)) > 0$, then Q is differentiable with derivative the quantile-density $q_X(p) = Q(p)' = \frac{1}{f(Q(p))}$. Then ψ is twice differentiable with Hessian matrix

$$\begin{bmatrix} \frac{\partial^2 \psi}{\partial m^2} & \frac{\partial^2 \psi}{\partial m \partial p} \\ \frac{\partial^2 \psi}{\partial m \partial p} & \frac{\partial^2 \psi}{\partial p^2} \end{bmatrix} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

with $a = 2p$, $b = 2m - Q(1-p)$, $c = \frac{2m}{f(Q(1-p))}$. The Hessian is positive-definite at the critical point (m_X, p_X) if $a > 0$ and $ac - b^2 > 0$. The latter condition writes

$$ac - b^2 = \frac{4p_X m_X}{f(Q(1-p_X))} - 4(m - Q(1-p_X))^2 > 0$$

With (m_X, p_X) satisfying (13), (14), the condition writes

$$4Q(1-p_X) \left(\frac{2p_X}{f(Q(1-p_X))} - Q(1-p_X) \right) > 0,$$

which is (15). □

Remark 3. 1. If $1 - p_X$ is in the range of F (in particular if F is continuous), then, (13) writes

$$F(m_X/2) = 1 - p_X,$$

and by Exercise 3.3 in Shorack (2000) p. 113, (14) writes as

$$m_X = \frac{\mathbb{E}[X \mathbf{1}_{X > m_X/2}]}{1 - F(m_X/2)} = \mathbb{E}[X | X > m_X/2].$$

This equation will be derived in the general case from the optimal quantization viewpoint, see Theorem 3.5 and Remark 4 below.

2. Setting $a := Q(1 - p_X)$, the sufficiency condition (15) writes also

$$af(a) < 2(1 - F(a)).$$

In view of the fact that $\mathbb{E}[X] = \int_0^\infty tf(t) = \int_0^\infty (1 - F(t))dt < \infty$, the latter condition has to be understood as a condition on the tail decrease of the density. It can also be expressed as a condition on the failure rate (or hazard function), as

$$h(a) := \frac{f(a)}{1 - F(a)} < \frac{2}{a}.$$

A general existence result will be given in Theorem 3.5, by quantization methods.

3.1.2 Examples

We illustrate the formulas obtained in Proposition 3.1 on the following examples:

Example 1 (Uniform distribution). For $X \sim U_{[0,a]}$, i.e. $Q(p) = ap$, (13) and (14) give $p_X = 2/3$ and $m_X = 2a/3$.

Example 2 (Exponential distribution). For $X \sim \text{Exp}(\lambda)$, with $\lambda > 0$, the memorylessness property yields that $\mathbb{E}[X | X > a] = a + \lambda^{-1}$. Hence, the optimal threshold is $a_X = \lambda^{-1} = \mathbb{E}[X]$ and one has $p_X = e^{-2} \approx 0.135$ and $m_X = \frac{2}{\lambda} = 2\mathbb{E}[X]$.

It is noteworthy that, in both examples, p_X is fixed and does not depend on the parameter (a , resp. λ) of the distribution.

Example 3 (Pareto distribution). *One considers the following one-parameter (version) of the Pareto Distribution, $X \sim Pa(\theta)$, defined by $P(X > x) = (1 + x)^{-\theta}$, for $x \geq 0$. We assume that $\theta > 2$, so that $\mathbb{E}[X^2] < \infty$. One has that*

$$\mathbb{E}[X|X > a] = \frac{\theta}{\theta - 1}(1 + a) - 1.$$

Using the characterisation (24), one finds that the optimal threshold is $a = \frac{1}{\theta - 2}$. Hence, (25) gives

$$m_X = \frac{2}{\theta - 2}, \quad p_X = \left(\frac{\theta - 2}{\theta - 1}\right)^\theta.$$

For some other distributions, one may not have a closed form expression.

3.2 A review of Optimal Quantization

Additional insight is gained by viewing the basic approach (11) as a constrained optimal quantization problem. We first summarize the basic facts and terminology of (unconstrained) optimal quantization theory. The following can be regarded as a quick introduction to the field.

Optimal quantization originates from the engineering and signal processing literature (see Lloyd (1982), Gersho and Gray (1992)). It aims at optimally discretizing a continuous (stationary) signal in view of its transmission. It was developed originally for analog-to-digital conversion, compression, pattern recognition.

Following Gersho and Gray (1992) and Graf and Luschgy (2000), a N -vector quantizer on $(\mathbb{R}^d, \|\cdot\|)$ is a mapping $T : \mathbb{R}^d \mapsto \{x_1, \dots, x_N\}$, where $\{x_1, \dots, x_N\}$ is a codebook of size N . Thus, associated with T is a partition $\{A_i\}$, with $A_i = \{x \in \mathbb{R}^d : T(x) = x_i\}$, of the input space so that a quantizer writes

$$T(x) = \sum_{i=1}^N x_i \mathbb{1}_{A_i}(x), \quad x \in \mathbb{R}^d$$

and is determined by the pairs $\{(x_i, A_i), i = 1, \dots, n\}$. An N -optimal quantizer for a distribution P^X is a N -quantizer which minimises the mean squared error (or distortion)

$$D(T; P^X) := \inf_T \mathbb{E}(X - T(X))^2.$$

Equivalently, it can be shown (Pollard (1982), Graf and Luschgy (2000)) that an optimal quantizer is a Monge map minimising the Wasserstein metric $W_2(P^X, P^Y)$ between P^X and P^Y , where P^Y is a discrete measure with N points.

It is known, see Gersho and Gray (1992), that given the codebook, the optimal partition is given by the Voronoi partition,

$$A_i := \{x \in \mathbb{R}^d : \|x - x_i\| \leq \|x - x_j\|, j \neq i\}.$$

Conversely, given a partition, the optimal codebook is given by the centers (centroids),

$$y_i = \mathbb{E}(X|X \in A_i).$$

These properties are the basis of Lloyd's iterative algorithm. This also explains why optimal quantizers are sometimes defined directly by their codebook and corresponding Voronoi partition, see e.g. Pagès (2018). As a consequence, the distortion writes as a sole function of the centers, as

$$D(T; P^X) = D_N(x_1, \dots, x_N) := \mathbb{E} \min_{x_i} \|X - x_i\|^2 \quad (16)$$

If the support of P^X has at least N elements, existence of an N -optimal quantizer follows from the fact that $(x_1, \dots, x_N) \rightarrow \sqrt{D_N(x_1, \dots, x_N)}$ is 1-Lipshitz and a non-trivial compactness argument, see e.g. Lemma 8 in Pollard (1982), Theorem 5.1 in Pagès (2018) or Theorem 4.12 in Graf and Luschgy (2000). In dimension one, a known sufficient condition for uniqueness of the N -optimal quantizer is that P^X be absolutely continuous with a log-concave density f . By Proposition 6.6 in Pagès (2018), if $x^N := (x_1, \dots, x_N)$ has pairwise distinct component and $P^X(\cup_{1 \leq i \leq N} \partial A_i) = 0$, then D_N is continuously differentiable with gradient $\nabla D_N = \left[\frac{\partial D_N}{\partial x_i} \right]$ with

$$\frac{\partial D_N}{\partial x_i}(x^N) = \mathbb{E}[2(x_i - X)\mathbb{1}_{X \in A_i}]. \quad (17)$$

These calculations are the basis for a stochastic gradient descent based algorithm known as Competitive Learning Vector Quantization, see Pagès (2018). In the one-dimensional case, explicit expressions of the gradient can be written in terms of the c.d.f. F and of the cumulative first moment function $K(x) := \mathbb{E}[X\mathbb{1}_{X \leq x}]$ as

$$\frac{\partial D_N}{\partial x_i}(x^N) = 2x_i \left[F(x_{i+\frac{1}{2}}) - F(x_{i-\frac{1}{2}}) \right] - 2 \left[K(x_{i+\frac{1}{2}}) - K(x_{i-\frac{1}{2}}) \right],$$

where $x_{i+\frac{1}{2}} = \frac{x_{i+1} + x_i}{2}$ corresponds to the boundaries of the Voronoi cells. Additional formulas for the Hessian are available (see Pagès (2018) p. 155). These formulas allow for a Newton-Raphson zero search procedure.

3.3 Existence and characterization of (m_X, p_X) by constrained optimal quantization

We can now tackle the study of (m_X, p_X) from the optimal quantization viewpoint. One thus introduces the constrained two-points quantizer with centers $\{x_0, x_1\} := \{0, m\}$, i.e. as a mapping $T : [0, \infty) \mapsto \{0, m\}$ with $T(x) = m\mathbb{1}_{x \geq a}$, where $a \geq 0$ is a threshold to determinate. Then, the optimal quantization problem with constrained knot at zero writes

$$\inf_{(a, m) \in [0, \infty)^2} \mathbb{E}[(X - T(X))^2].$$

By the results of the previous section (see (16)), one already knows that $a = m/2$, i.e. that the Voronoi regions write $A_0 = \{x : 0 \leq x \leq m/2\}$ and $A_1 = \{x : x \geq m/2\}$. As a consequence, and in view of (16) the distortion/objective function writes as a sole function of the magnitude m , as

$$L(m) := \mathbb{E}[X^2 \wedge (X - m)^2]. \quad (18)$$

We first study the differentiability properties of the objective function (18) in the next Section.

3.3.1 B-differentiability properties of the objective function

For the general N -points quantization problem, the square root of the distortion function is 1-Lipshitz, see Pagès (2018) p. 136. Hence, it has a derivative a.e. by Rademacher's Theorem. Therefore, L is differentiable a.e. In addition, since the integrand $H : x \rightarrow X^2 \wedge (X - x)^2$ is piecewise differentiable, it is easy to show that it has directional derivatives everywhere. A convenient tool in the setting of optimisation of piecewise smooth functions is the concept of Bouligand derivative (B-derivative), see Robinson (1987) and Scholtes (2012). It drops the requirement of linearity of the differential, represents a first-order approximation and allows to have a single-valued notion of differential. We give below a simplified definition for functions of one variable.

Definition 3.2. *A function $f : \mathbb{R} \rightarrow \mathbb{R}$ is Bouligand differentiable (B-differentiable) at x_0 if there exists a positive homogeneous function $\nabla^B f(x_0) : \mathbb{R} \rightarrow \mathbb{R}$ s.t.*

$$f(x_0 + v) = f(x_0) + \nabla^B f(x_0)(v) + o(v), \quad \forall v \in \mathbb{R}. \quad (19)$$

For the study of the B-differentiability of the objective function L , we first give a lemma on the differentiability of its integrand.

Lemma 3.3 (B-derivative of a min function). *Let $H(x) = \min(h_1(x), h_2(x))$, where $h_1, h_2 : \mathbb{R} \rightarrow \mathbb{R}$ are differentiable functions. Then H is B-differentiable, with B-derivative given by*

- i) *if x is s.t. $h_1(x) < h_2(x)$, then $\nabla^B H(x)(v) = h'_1(x)v$, and conversely, if $h_1(x) > h_2(x)$, then $\nabla^B H(x)(v) = h'_2(x)v$.*
- ii) *if x is s.t. $H(x) = h_1(x) = h_2(x)$, then $\nabla^B H(x)(v) = \min(h'_1(x)v, h'_2(x)v)$.*

Proof. i) Assume w.l.o.g. that $h_1(x) < h_2(x)$, so that $H(x) = h_1(x)$. Let us show that $H(x) = h_1(x)$ remains true on a neighborhood of x . Set $d = h_2(x) - h_1(x) > 0$. Let $0 < \varepsilon_1 < d$, $0 < \varepsilon_2 < d - \varepsilon_1$. By continuity of h_1 and h_2 at x , there exists $\delta > 0$ s.t. $\forall |v| < \delta$,

$$h_1(x + v) \leq h_1(x) + \varepsilon_1 \quad (20)$$

$$h_2(x + v) \geq h_2(x) - \varepsilon_2 \quad (21)$$

$h_1(x) + \varepsilon_1 = h_2(x) + \varepsilon_1 - d$, therefore (20) and (21) give

$$h_1(x + v) \leq h_2(x) + \varepsilon_1 - d < h_2(x) - \varepsilon_2 \leq h_2(x + v).$$

Hence, $H(x + v) = h_1(x + v)$ for all $|v| < \delta$. Thus, the Bouligand derivative of H at x reduces to the classical derivative $h'_1(x)$ of h_1 .

ii) One has

$$\begin{aligned} h_1(x + v) &= h_1(x) + h'_1(x)v + o(v) \\ h_2(x + v) &= h_2(x) + h'_2(x)v + o(v) \end{aligned} \quad (22)$$

Since $H(x) = h_1(x) = h_2(x)$,

$$H(x + v) - H(x) = \min(h_1(x + v) - h_1(x), h_2(x + v) - h_2(x)).$$

We then use the inequality

$$|\min(a, b) - \min(c, d)| \leq \max(|a - c|, |b - d|),$$

applied to $a = h_1(x+v) - h_1(x)$, $b = h_2(x+v) - h_2(x)$, $c = h'_1(x)v$, $d = h'_2(x)v$, to deduce that

$$|H(x+v) - H(x) - \min(h'_1(x)v, h'_2(x)v)| \leq \max(|o(v)|, |o(v)|) = |v| \max(|o(1)|, |o(1)|),$$

viz. H is B-differentiable at x , with B-derivative

$$\nabla^B H(x)(v) = \min(h'_1(x)v, h'_2(x)v),$$

as the latter expression is positively homogeneous. \square

Corollary 3.4 (B-Differentiability of the objective function). *L is B-differentiable on $x \geq 0$, with B-derivative given by*

$$\nabla^B L(0)(v) = -2\mathbb{E}[X]v\mathbf{1}_{v>0},$$

and, for $x > 0$,

$$\nabla^B L(x)(v) = 2v\mathbb{E}[(x-X)\mathbf{1}_{X>x/2}] + P(X=x/2)\min(0, xv). \quad (23)$$

Proof. Applying Lemma 3.3 to the integrand function $H(x) = X^2 \wedge (X-x)^2$, i.e. with $h_1(x) = X$, $h_2(x) = (X-x)^2$, gives the B-derivative of the integrand function: almost surely, for $v \in \mathbb{R}$,

$$\nabla^B H(x)(v) = \begin{cases} 0 & x < 0 \\ \min(0, -2Xv) & x = 0 \\ 2(x-X)v & 0 < x < 2X \\ \min(0, 2Xv) & x = 2X \\ 0 & x > 2X \end{cases}$$

In addition, the second order terms in the Taylor expansions (22) of h_1 and h_2 (which are exact) do not depend on X , so that reasoning as in 3.3, it is easy to see that one gets a remainder term in (19) for H which does not depend on X , with at most linear growth:

$$H(x+v) = H(x) + \nabla^B H(x)(v) + v\epsilon(v),$$

a.s., with $|\epsilon(v)| \leq |v|$. Therefore, if one sets

$$\nabla^B L(x)(v) := \mathbb{E}[\nabla^B H(x)(v)],$$

then, by linearity,

$$\frac{|L(x+v) - L(x) - \nabla^B L(x)(v)|}{|v|} \leq |v| \rightarrow 0,$$

as $|v| \rightarrow 0$. By integration,

$$\nabla^B L(0)(v) = \begin{cases} -2\mathbb{E}[X]v & v > 0 \\ 0 & v \leq 0 \end{cases}$$

and for $x > 0$,

$$\begin{aligned} \nabla^B L(x)(v) &= \int (2(x-t)\mathbf{1}_{0<x<2t} + \min(0, 2tv)\mathbf{1}_{x=2t}) P^X(dt) \\ &= \mathbb{E}[2(x-X)\mathbf{1}_{X>x/2}]v + P(X=x/2)\min(0, xv). \end{aligned}$$

$v \rightarrow \nabla^B L(x)(v)$ is positively homogeneous, for $x \geq 0$. Hence, L is B-differentiable. \square

3.3.2 Main existence and characterization result

With these tools, we can now give the main existence and characterization result of the optimal magnitude-propensity pair.

Theorem 3.5. *i) If $\mathbb{E}[X^2] < \infty$ and the support of P^X contains at least two points, then there exists a magnitude-propensity pair (p_X, m_X) minimizing (11).*

ii) An optimal magnitude-propensity pair (p_X, m_X) is characterized by solving for a the equation

$$2a = \mathbb{E}[X|X > a], \quad a > 0, \quad (24)$$

and then setting

$$m_X = 2a, \quad p_X = P(X > a). \quad (25)$$

In addition, $P^X(\{a\}) = 0$.

Proof. i) By the results of the Section 3.2, the objective function L of (18) is locally Lipschitz hence continuous and its lower level sets are compact, see e.g. Lemma 8 in Pollard (1982). Hence, the general existence result follows from Weierstrass theorem. See also Exercise 3 p. 139 in Pagès (2018).

ii) Since $\nabla^B L(\cdot)(\cdot)$ gives a first order approximation of L , a necessary condition for m_X to be a minimizer is that $\nabla^B L(m_X)(v) \geq 0, \forall v \in \mathbb{R}$. For $m_X = 0$, $\nabla^B L(0)(v) = -2\mathbb{E}[X] < 0$ for $v > 0$, therefore 0 cannot be a critical point. For $m_X > 0$ and $v > 0$, the first order condition writes

$$\mathbb{E}[(m_X - X)\mathbf{1}_{X > m_X/2}] = 0. \quad (26)$$

For $m_X > 0, v < 0$, it reduces to

$$2\mathbb{E}[(m_X - X)\mathbf{1}_{X > m_X/2}] + m_X P(X = m_X/2) = 0. \quad (27)$$

Plugging (26) into (27) implies that $P(X = m_X/2) = 0$ and yields (24), upon reparametrizing by $a = m_X/2$. □

Remark 4.

i) We hopefully obtain the same formula between (24) of Theorem 3.5 and (13), (14) of Proposition 3.1. The quantization approach allows to obtain a general existence result.

ii) P^X does not charge the optimal threshold $a = m_X/2$, even if P^X is not absolutely continuous. By (23), this imply that $\nabla^B L(m_X)(v)$ is linear in v , that is to say, L is differentiable (in the classical sense) at the optimal m_X , with

$$L'(m_X) = 2\mathbb{E}[(m_X - X)\mathbf{1}_{X > m_X/2}] = 2\mathbb{E}[(m_X - X)\mathbf{1}_{X \geq m_X/2}].$$

3.3.3 Uniqueness

The matter of uniqueness is notoriously a difficult topic in optimal quantization, see Graf and Luschgy (2000), Pagès (2018). We provide below in Theorem 3.6 a set of sufficiency conditions for the uniqueness of the magnitude-propensity pair. The main condition is the log-concavity of $x \rightarrow x^3 f(x)$, which is fulfilled when the density f is itself log-concave. It is a condition similar to the classical condition of Trushkin (1982) and Kieffer (1983) for the uniqueness in (unconstrained) optimal quantization.

Theorem 3.6. Let P^X be an absolutely continuous distribution on \mathbb{R}_+ , with density f , $0 \in \text{supp}(P^X)$ and $\mathbb{E}[X^2] < \infty$. Assume

- i) $\tilde{f} : x \rightarrow x^3 f(x)$ is strictly log-concave on \mathbb{R}_+ ;
- ii) $\lim_{x \rightarrow 0} x f(x) = 0$.

Then, the optimal magnitude-propensity pair is unique.

Proof. Prior to the proof, let us make the following elementary remarks. Being log-concave, the function $\tilde{f} : x \rightarrow x^3 f(x)$ has a limit in $[0, \infty)$ as $x \rightarrow 0$. Thus, we may set $\tilde{f}(0) = \lim_{x \rightarrow 0} x^3 f(x) = 0$, by assumption ii). Consequently, $I := \{x > 0 : \tilde{f}(x) > 0\}$ is an interval of \mathbb{R}_+ of the form $(0, b]$ or $(0, b)$, with $b \in (0, \infty]$, since $0 \in \text{supp}(P^X)$, and $f(x) = \tilde{f}(x)/x^3$ is continuous on $(0, b)$. Moreover, still by log-concavity, \tilde{f} has a left-handed limit $\tilde{f}(b-)$ at b . Finally, f is continuous on $(0, b)$, with left-handed limit $f(b-) \geq 0$.

The optimal magnitude m_X is s.t. $y_* = m_X/2$ is a zero of the function

$$\Phi(y) = 2y\bar{F}(y) - \bar{K}(y), \quad (28)$$

where we have set $\bar{F}(y) = \int_y^b f(t)dt$ and $\bar{K}(y) = \int_y^b t f(t)dt$. Moreover, since \tilde{f} is log-concave, $\ln f$ has right and left-handed derivatives on $(0, b)$, and, by monotonicity, these one-sided derivatives have limits as $x \downarrow 0$ and $x \uparrow b$. Therefore, Φ' can be continuously defined on $y \in [0, b] \cap \mathbb{R}^+$, as

$$\Phi'(y) = 2\bar{F}'(y) - yf(y). \quad (29)$$

On $(0, b)$, Φ' has right and left-handed derivatives, so let us define Φ'_r as the right-handed derivative

$$\begin{aligned} \Phi''_r(y) &:= -3f(y) - yf'_r(y) = -yf(y) \left(\frac{3}{y} + \frac{f'_r}{f}(y) \right) \\ &= -yf(y)\zeta(y), \end{aligned} \quad (30)$$

where

$$\zeta(y) := 3/y + f'_r(y)/f(y) \quad (31)$$

with f'_r denoting the right-handed derivative of f . By assumption, $x \rightarrow x^3 f(x)$ is strictly log-concave, therefore ζ is decreasing.

Let us show that there exists a unique $y_1 \in (0, b)$ s.t. $\Phi'(y_1) = 0$. For that purpose, consider the following cases:

Case i) Assume that $\lim_{y \rightarrow b} \zeta(y) = \ell \geq 0$. Then, by strict log-concavity, $\zeta(y) > \ell \geq 0$ on I , i.e. $\Phi'' < 0$ and Φ' decreasing on $[0, b]$. By assumption ii), $\Phi'(0) = 2\bar{F}(0) - \lim_{y \rightarrow 0} yf(y) = 2 > 0$.

Case ia) if $b = \infty$, then $\lim_{y \rightarrow \infty} \Phi'(y) = -\lim_{y \rightarrow \infty} yf(y) = 0$ since $\mathbb{E}[X] < \infty$. Thus $\Phi' > 0$ and Φ is increasing on $(0, \infty)$. But since $\Phi(0) = -\mathbb{E}[X] < 0$ and $\lim_{y \rightarrow \infty} \Phi(y) = 0$, Φ cannot have a zero in $(0, \infty)$, which contradicts the general existence Theorem 3.5. Thus, necessarily, $b < \infty$.

Case ib) if $b < \infty$, then $\lim_{y \rightarrow b} \Phi'(y) = -bf(b-) \leq 0$. If $f(b-) = 0$, then $\Phi' > 0$ on $(0, b)$, so that Φ is increasing on $[0, b]$ from $-\mathbb{E}[X]$ up to 0, which contradicts the existence of $y_* = m_X/2 \in [0, b/2)$ such that $\Phi(y_*) = 0$. Therefore, $f(b-) > 0$, $\lim_{y \rightarrow b} \Phi'(y) < 0$ and there exists a unique $y_1 \in (0, b)$ s.t. $\Phi'(y_1) = 0$.

Case ii) Assume that $\lim_{y \rightarrow b} \zeta(y) = \ell < 0$.

Case iia) If $\lim_{y \rightarrow 0} \zeta(y) = \eta \leq 0$. Then, since ζ is decreasing, $\zeta < 0$ on $(0, b)$, hence $\Phi_r'' > 0$, so that Φ' is increasing on $(0, b)$. But this is clearly impossible, since $\Phi'(0) = 2$ and $\lim_{y \rightarrow b} \Phi'(y) = 0$, resp. $= -bf(b-) \leq 0$, for $b = \infty$, resp. $b < \infty$. Thus, necessarily $\lim_{y \rightarrow 0} \zeta(y) = \eta > 0$.

Case iib) If $\lim_{y \rightarrow 0} \zeta(y) = \eta > 0$. Then, since ζ is decreasing, there exists a unique $y_0 \in I$ s.t. $\Phi_r'' < 0$ on $(0, y_0)$, and $\Phi_r'' > 0$ on (y_0, b) .

Assume that $\Phi'(y_0) \geq 0$. Then, $\Phi' > 0$ on $(0, b) \setminus \{y_0\}$ and Φ is increasing on $[0, b]$. By the general existence Theorem 3.5, there exists some $y^* = m_X/2 \in (0, b)$ s.t. $\Phi(y^*) = 0$. Yet, $\lim_{y \rightarrow \infty} \Phi(y) = 0$ since $\mathbb{E}[X] < \infty$, which is a contradiction. Therefore, $\Phi'(y_0) < 0$. With the fact that $\Phi'(0) = 2 > 0$ and Φ' is decreasing on $(0, y_0)$, this implies that there exists a unique $y_1 \in (0, y_0)$ s.t. $\Phi'(y_1) = 0$. On (y_0, b) , Φ' cannot have any zeros, since Φ' is increasing, $\Phi'(y_0) < 0$ and either $\Phi'(b-) = -bf(b-) < 0$ if $b < \infty$, or $\lim_{y \rightarrow \infty} \Phi'(y) = 0$ if $b = \infty$, since $\mathbb{E}[X] < \infty$.

So, in both cases, Φ is increasing on $(0, y_1)$ and decreasing on (y_1, b) . Moreover $\Phi(0) = -\mathbb{E}[X] < 0$ and $\lim_{y \rightarrow b} \Phi(y) = 0$, and the general existence Theorem 3.5 ensures that Φ has at least a zero in $(0, b/2)$. Hence, necessarily $\Phi(y_1) \geq 0$ and Φ has at most one zero, located in $(0, y_1)$. \square

In particular, if f is itself log-concave, then the conditions of Theorem 3.6 are fulfilled, as shown in the next corollary.

Corollary 3.7. *If f is a log-concave density on $(0, b)$, then the assumptions i) ii) of Theorem 3.6 hold.*

Proof. i) $x \rightarrow 3/x$ is decreasing and f_r'/f is non-increasing, therefore ζ is decreasing, i.e. $\tilde{f} : x \rightarrow x^3 f(x)$ is strictly log-concave.

ii) Let $y_0 > 0$ s.t. $f(y_0) > 0$. For $0 < y < y_0 < b$, one has, by log-concavity, that

$$\begin{aligned} \ln f\left(\frac{y+y_0}{2}\right) &\geq \frac{1}{2} \ln f(y) + \frac{1}{2} \ln f(y_0) \\ \iff \ln f(y) &\leq 2 \ln f\left(\frac{y+y_0}{2}\right) - \ln f(y_0) \\ \iff 0 \leq f(y) &\leq f^2\left(\frac{y+y_0}{2}\right) / f(y_0) \\ \iff 0 \leq yf(y) &\leq yf^2\left(\frac{y+y_0}{2}\right) / f(y_0), \end{aligned}$$

which implies $\lim_{y \rightarrow 0} yf(y) = 0$ by continuity. (Note that the statement is trivial if $\lim_{y \rightarrow 0} f(y) < \infty$.) \square

Remark 5. *Examples of distributions satisfying the hypotheses of Theorem 3.6, but not those of Corollary 3.7 (in particular, whose density is not log-concave) are the power distributions*

$$f(x) = (a+1)x^a \mathbb{1}_{(0,1)}(x), \quad -1 < a < 0,$$

and the Gamma distributions

$$f(x) = \frac{x^{a-1}}{\Gamma(a)} e^{-x} \mathbb{1}_{x>0}, \quad 0 < a < 1.$$

Remark 6. *If, in addition to being log-concave, the density is non-increasing, a simpler proof of Theorem 3.6 can be given as follows: Let \tilde{X} be the symmetrized version of X , with density $\tilde{f}(x) = \frac{1}{2}f(|x|)$. Let $\lambda \in [0, 1]$, $x, y \in \mathbb{R}$. Since $\ln f$ is non-increasing and concave,*

$$\ln f(|\lambda x + (1 - \lambda)y|) \geq \ln f(\lambda|x| + (1 - \lambda)|y|) \geq \lambda \ln f(|x|) + (1 - \lambda)\ln f(|y|),$$

viz. \tilde{f} is log-concave. Therefore, by the existence and uniqueness results of Trushkin (1982), Kieffer (1983), there exists a unique three-points stationary quantizer of \tilde{X} .

The optimal magnitude m_X is obtained by deriving the optimal two-points quantizer of X with a point constrained to be zero. m_X must verify the stationary condition

$$\mathbb{E}[(m_X - X)\mathbb{1}_{X \geq m_X/2}] = 0. \quad (32)$$

On the other hand, a three-points (unconstrained) quantizer (x_1, x_2, x_3) for \tilde{X} must satisfy the stationary condition $\mathbb{E}[(x - \tilde{X})\mathbb{1}_{A_i}(\tilde{X})] = 0$. Let us show that that $(-m_X, 0, m_X)$ is a three-points stationary quantizer for \tilde{X} . Indeed, for $x_1 = -m_X$, the stationary condition writes $\int(-m_X - t)\mathbb{1}_{t \leq m_X/2}\tilde{f}(t)dt = 0$, which gives (32) by the change of variable $x = -t$ and the symmetry of \tilde{f} . For $x_3 = m_X$, the stationary condition is also (32) since $m_X > 0$. For $x_2 = 0$, it writes $\int_{-m_X/2}^{m_X/2} x\tilde{f}(x)dx = 0$, which is automatically satisfied since \tilde{f} is even. Therefore, $(-m_X, 0, m_X)$ is the unique three points stationary quantizer for \tilde{X} . Thus m_X is unique.

3.4 Discussion and properties

The optimal magnitude-propensity pair (m_X, p_X) obtained by (25) or (13), (14) has several interesting characteristics. First, from (13), resp. (14), it is clear that the magnitude m_X can be interpreted either as (twice) a Value-at-Risk $VaR_\alpha(X)$, resp. as a Conditional Value-at-Risk $CVaR_\alpha(X)$, for a special value of α :

$$m_X = CVaR_{p_X}(X) = 2VaR_{1-p_X}(X). \quad (33)$$

It is thus comforting that one obtains, for the magnitude effect, a quantity akin to the classical and well-used univariate risk measures. In addition, such an interpretation gives an answer to the problem mentioned in Remark 2 about the “right” choice of α for parametrized risk measures like $VaR_\alpha(X)$ or $CVaR_\alpha(X)$: the corresponding α is determined by p_X , hence by the distribution P^X . It is no longer a subjective choice dependent on the user.

Second, the magnitude-propensity paradigm m_X, p_X , by quantifying risk on a bivariate scale, is fundamentally different from the traditional approach to measuring risk by a univariate coherent risk measure ρ . Hence, it may not make sense to look for properties similar to those of coherent risk measures. However, one has the following noteworthy properties for the magnitude m_X .

Proposition 3.8 (positive homogeneity/scaling of the magnitude). *For $a > 0$, $m_{aX} = am_X$, and $p_{aX} = p_X$*

Proof. From the expression (16) of the distortion as a function of the centers,

$$W_2^2(P^{aX}, P^Y) = \mathbb{E}[\min\{aX^2, (aX - m)^2\}] = a^2 \mathbb{E}[\min\{X^2, (X - m/a)^2\}].$$

Therefore $\frac{m_{aX}}{a} = m_X$, and $p_{aX} = p_X$. □

Recall the definition of the convex order: $X \leq_{cx} Y$ if $\mathbb{E}\phi(X) \leq \mathbb{E}\phi(Y)$, for all integrable, convex functions ϕ , see e.g. Rüschemdorf (2013). One has:

Proposition 3.9 (Monotonicity of the magnitude). *If $X \leq_{cx} Y$, then $m_X \leq m_Y$.*

Proof. By e.g. Proposition 1 in Puccetti (2013), $X \leq_{cx} Y$ if and only if $\mathbb{E}[X|X > a] \leq \mathbb{E}[Y|Y > a]$, for all $a \geq 0$. Hence, the curve $a \mapsto \mathbb{E}[X|X > a]$ is below the curve $a \mapsto \mathbb{E}[Y|Y > a]$. Thus, the optimal threshold a_X for X , satisfying (24), is lower than the corresponding optimal threshold a_Y for Y . Therefore, $m_X \leq m_Y$. \square

Eventually, the magnitude effect will be larger than the average of X , which seems a desirable property from the point of view of the classical premium calculation principles Bühlmann (1996):

Proposition 3.10. *If P^X is absolutely continuous, then $m_X \geq EX$.*

Proof. Let $\bar{F}(x) = P(X \geq x)$, $\bar{K}(x) = \mathbb{E}[X\mathbf{1}_{X \geq x}] = \int_{(x, \infty)} tf(t)dt$ and $\beta(x) = \bar{K}(x)/\bar{F}(x)$. Then,

$$\beta'(x) = \frac{-xf(x)\bar{F}(x) + f(x)\bar{K}(x)}{\bar{F}^2(x)} = \frac{f(x)}{\bar{F}(x)} \left(\frac{\bar{K}(x)}{\bar{F}(x)} - x \right) \geq 0$$

as $\bar{K}(x) \geq x\bar{F}(x)$. Therefore, β is non-decreasing, hence $\beta(x) \geq \beta(0) = \mathbb{E}[X]$. Therefore, the solution m_X of the equation $x = \beta(x/2)$ satisfy $m_X \geq \mathbb{E}[X]$. \square

4 Magnitude-propensity risk comparison: numerical illustrations and empirical aspects

4.1 Risk comparison with magnitude-propensity plots

A fundamental objective of risk analysis is the comparison of risks. The proposed approach allows to compare risks on both the magnitude and propensity scales, by displaying a point of coordinates (m_X, p_X) on magnitude and propensity axes. Figure 2 illustrates the resulting magnitude-propensity plot, for the uniform, exponential and Pareto distributions of Examples 1-3. Both the uniform and exponential distribution have constant propensity along the parameter, and a magnitude linearly increasing with the mean of the distribution. The blue dots show the (m_X, p_X) points for a distribution with mean $\mathbb{E}[X] = 1, 2, 5$, and 10: the uniform distribution has a higher propensity, with a lesser magnitude effect. For the Pareto distribution, the red circles show the (m_X, p_X) points with parameter value $\theta = 2.1, 2.5, 5, 10$: the tail effect is reflected in the behavior of the magnitude-propensity pair. For a heavy-tailed distribution (θ close to 2), one has a large magnitude with a very small propensity, while for a short-tailed distribution (θ large), the magnitude effect remain limited but with a larger propensity. This behaviour is in accordance with what could be expected from intuition.

4.2 Distributions with no closed form expressions for (m_X, p_X)

For some more complicated distributions, one may not have a closed form expression of the (m_X, p_X) as in Examples 1-3. However, one can use the computation capabilities of, e.g., Mathematica Wolfram Research, Inc. (2020) to obtain a closed-form

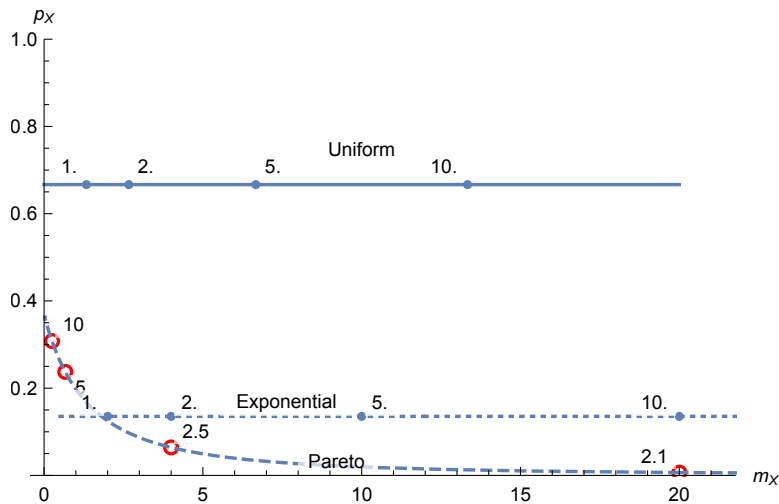


Figure 2: Magnitude-propensity plots for the uniform $U_{[0,a]}$ distribution (solid line), Exponential $Exp(\lambda)$ (dotted) and Pareto $Pa(\theta)$ (dashed), for varying values of the parameter a, λ, θ .

expression of the $\mathbb{E}[X|X > a]$ and then find numerically the root of (24). We illustrate the procedure for the Gamma and Weibull distributions, two distributions frequently encountered in insurance mathematics.

Example 4 (Gamma distribution). For $X \sim \Gamma(\alpha, \beta)$, its density writes $f(x) = x^{\alpha-1} e^{-(x/\beta)}$. One has

$$\mathbb{E}[X|X > a] = \beta \frac{\Gamma(1 + \alpha, a/\beta)}{\Gamma(\alpha, a/\beta)}$$

where Γ is the incomplete Gamma function. We take a family of $\Gamma(\alpha, 2)$ distribution, with shape parameter α varying from 0.1 to 2.9 by stepsize of 0.2. We solve (24) numerically with initial point $\mathbb{E}[X]$. The resulting magnitude-propensity plot is displayed in Figure 4.

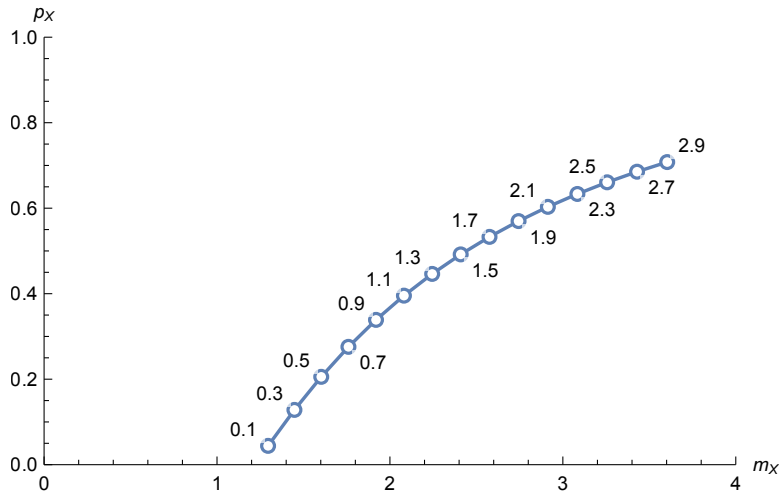


Figure 3: Magnitude-propensity plot for Gamma distribution $X \sim \Gamma(\alpha, 2)$, for α varying from 0.1 to 2.9. The corresponding mean $\mathbb{E}[X]$ is indicated near the circled (m_X, p_X) points.

The coordinates of the circled points indicate the magnitude-propensity (m_X, p_X) values and are labeled with the expectation $\mathbb{E}[X]$. Here, it is interesting that both m_X and p_X are increasing with the mean parameter, contrary to the Pareto case of Figure 2. This corresponds to intuition: let us recall that as the shape parameter α increase, the density $f(x)$ is shifted to the right and shrinks toward zero for small x values, see Figure 4. Hence, both the magnitude and propensity coding the risk borne by X will increase.

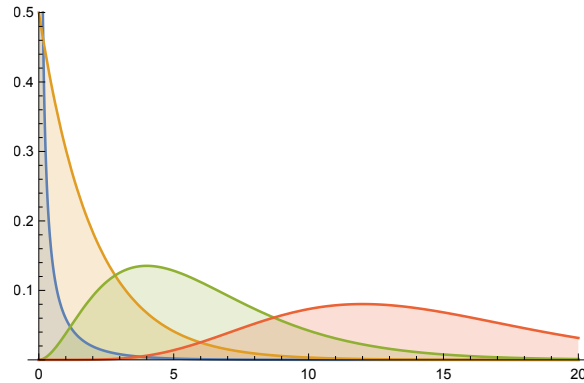


Figure 4: Probability density functions f for Gamma distributions $\Gamma(\alpha, 2)$, with $\alpha = 0.1$ (blue), 1 (orange), 3 (green), 7 (red).

Example 5 (Weibull distribution). For $X \sim W(\alpha, \beta)$, its density writes $f(x) = x^{\alpha-1} e^{-(x/\beta)^\alpha}$. One has

$$\mathbb{E}[X|X > a] = \beta e^{a^\alpha \beta^{-\alpha}} \Gamma(1 + 1/\alpha, a^\alpha \beta^{-\alpha})$$

where Γ is the incomplete Gamma function. The effect of shifting the parameters is illustrated in Figure 5 below.

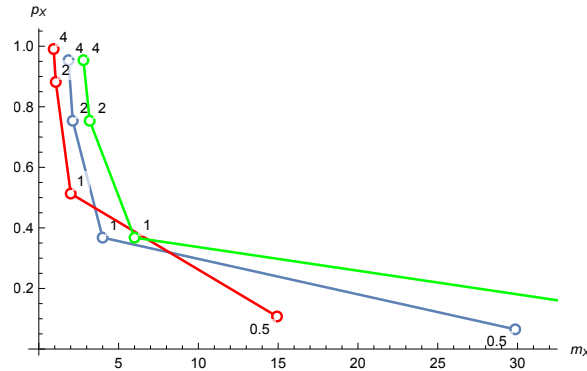


Figure 5: Magnitude-propensity plot for a Weibull $W(\alpha, \beta)$ distribution for $\beta = 2$ (blue), $\beta = 1.5$ (red), $\beta = 3$ (green). The magnitude-propensity (m_X, p_X) points are labeled by the values of $\alpha = 0.5, 1, 2, 4$.

4.3 Empirical computations and illustrations

The characterization (24) shows that the optimal threshold is a fixed point of the function

$$a \mapsto \frac{\mathbb{E}[X|X > a]}{2}. \quad (34)$$

For empirical data, i.e. when the distribution of X is unknown but one has a sample of realizations of X , one can replace the unknown expectation in (34) by a sample estimator. For the empirical measure, one obtains an estimate of (m_X, p_X) as a fixed point of

$$a \mapsto \frac{\sum_{i=1}^n X_i \mathbb{1}_{X_i > a}}{2 \sum_{i=1}^n \mathbb{1}_{X_i > a}}.$$

A Mathematica code is given in the supplementary appendix. The latter would correspond to Lloyd’s algorithm in the unconstrained quantization problem. See Pagès (2018) for discussion of numerical optimal quantization. Alternatively, one could seek directly for a (global) minimizer of the objective function L of (18), using a generic minimizer program (e.g. the command “NMinimize” in Wolfram Research, Inc. (2020)).

These approaches are illustrated on a synthetic and a real dataset. For the synthetic data set, we simulated a sample of 100 i.i.d. Uniform on $[0, 1]$ random variables. From Example 1, we know that we should obtain $(m_X, p_X) = (2/3, 2/3)$. Lloyd’s algorithm on the sample gives

$$(m_X, p_X) \approx (0.67, 0.65).$$

NMinimize, which finds a global minimum, also finds the same $m_X \approx 0.67$.

For the real data set, we take the US Hurricane Losses data, which is an example data set in Wolfram Research, Inc. (2020). It reports the thirty most destructive hurricanes in the U.S., from 1949 to 1999, see Figure 4.3.

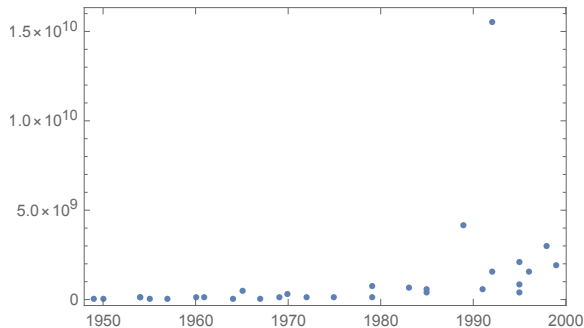


Figure 6: US Hurricane Losses data (Reported losses) in the U.S. 1949 – 1999. Available at <https://datarepository.wolframcloud.com/resources/US-Hurricane-Loss>.

We standardized the original data by dividing each entry by 10^6 . There is one clear outlier at 15500 (the famed “Andrew” hurricane). It is noteworthy that Lloyd’s algorithm on the full dataset give

$$(m_X, p_X) \approx (15500, 0.0333),$$

i.e. the outlier value as magnitude, with propensity one over the number of sample points ($1/30$): the outlier has a dwarfing effect on the whole data set and this is reflected on the magnitude-propensity measure. If the outlier value is removed, one gets

$$(m_X, p_X) \approx (2401.67, 0.206897),$$

which better reflects the magnitude and propensity effects of the data set.

5 Perspective: towards general risk quantization

In order to make the presentation clear, the discussion so far has been directed towards

- i) univariate risks $X \in [0, \infty[$;
- ii) using the simplest way to quantify risk on the magnitude and propensity scale, i.e. on the basic idea (11) with a constrained two-points quantizer $\{0, m_X\}$.

As a constrained optimization problem, the proposed approach naturally suggests to consider several variants and extensions. To stimulate further research on the topic, we conclude the paper with a brief sketch below of some of these possible variants and extensions.

5.1 Variant: three-points quantification

One can refine the quantized proxy by looking for a quantization to a discrete distribution with more than two points. For example, one can refine one’s measure of risk into a classification between a “moderate” and a “large” risk (or, say, “low” risk and “tail” risk), by using a three points discrete measure, $P^Y = (1 - p_1 - p_2)\delta_0 + p_1\delta_{m_1} + p_2\delta_{m_2}$, with $m_1 < m_2$. With a such three points discrete measure, one can encode and quantify in the magnitude and propensity scale both moderate, resp. large risk, with (m_1, p_1) , resp. (m_2, p_2) .

Figure 5.1 illustrates the proposed variant on a simulated dataset of size 1000 sampled from a standard log-normal distribution. The theoretical density function (blue curve) together with the histogram of the data is plotted along with the three quantization points (red circles) of respective magnitudes 0, 2.165, 8.200, with corresponding propensities 0.539, 0.393, 0.068: one has a “moderate” risk $(m_1, p_1) = (2.165, 0.393)$ of relatively low magnitude and high propensity, and a tail risk $(m_2, p_2) = (8.2, 0.068)$ of large magnitude but low propensity. This example illustrates the duality of risk exposed in Section 2 and the relevance of the proposed paradigm in distinguishing both phenomena and quantifying them.

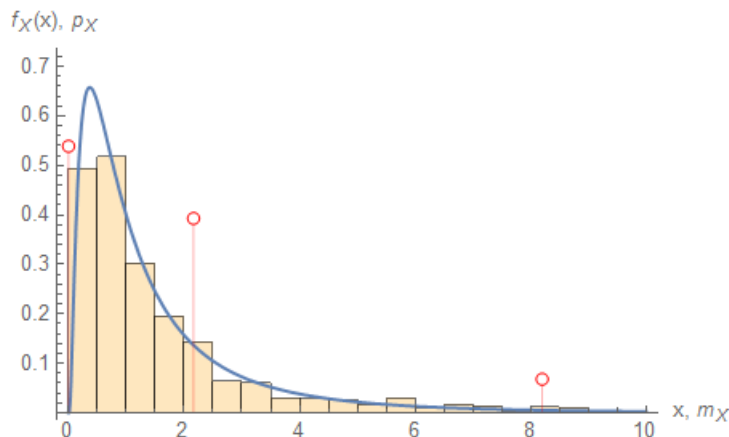


Figure 7: Three points risk quantization from a sample of size 1000 sampled from a standard log-normal distribution. Density function $f_X(x)$ (blue curve). Risk quantization points (m_X, p_X) (red circles).

Such a simultaneous quantization of the “mild” and “extreme” magnitude-propensity parts of an insurance risk X could reveal to be very valuable in applications. For example, in Reinsurance (see e.g. Albrecher, Beirlant, and Teugels (2017)), the insurer and reinsurer have to agree on an optimal risk sharing policy, with the insurer usually ceding to the reinsurer the “extreme” part of the risk (the one with high magnitude and low propensity), while keeping and managing the “mild” part. The proposed 2–levels risk quantization approach seems to be particularly well-suited for identifying and quantifying these parts. Also, for capital adequacy, the “tail” magnitude-propensity risk measure (m_2, p_2) could serve as a basis from which one can compute a sufficient safety margin against adverse extreme risks.

Eventually, let us remark that one can of course increase the number of points in the quantization proxy to more than three points. From an approximation viewpoint, the more points, the better is the quantization of the distribution. However, the more points one uses, the more complicated the framework becomes, as the number of magnitudes and propensities measures increase, rendering the task of comparing and summarizing risks less straightforward. Hence, there is a balance to reach between simplicity of the framework and accuracy of approximation.

5.2 Extension to real-valued risk

So far, we have restricted on non-negative risk, mainly for clarity and simplicity of presentation. There are several variants which can be proposed to accommodate for a real-valued risk $X \in \mathbb{R}$ (so that, in the insurance convention where X stands for a loss, $X < 0$ stands for a gain).

1. First, one can apply the previous framework to the conditional distribution of the positive part $X|X \geq 0$ or to $X^+ = \max(X, 0)$. This makes sense if the gains do not offset the losses in the application considered, so that one can restrict the measure of risk to the losses only.
2. Second, one can look for three-points risk quantization to account for the negative part of the distribution. For example, one can apply the framework of (11) with

$$P^Y(\cdot) = p_- \delta_{m_-}(\cdot) + (1 - p_+ - p_-) \delta_0(\cdot) + p_+ \delta_{m_+}(\cdot), \quad (35)$$

with $m_- \leq 0 \leq m_+$. The latter constraint allows for an easy interpretation of (m_+, p_+) , resp. (m_-, p_-) , as the magnitude-propensity measure of the loss, resp. gain, part of the distribution.

Another possibility is to impose a quantification point at the mean, instead of zero, i.e. to use as a target distribution

$$P^Y(\cdot) = p_- \delta_{m_-}(\cdot) + (1 - p_+ - p_-) \delta_{\mathbb{E}[X]}(\cdot) + p_+ \delta_{m_+}(\cdot), \quad (36)$$

with $m_- \leq \mathbb{E}[X] \leq m_+$. Here, the interpretation is also straightforward: the proxy Y takes a “large” value m_+ with likelihood p_+ , a “small” value m_- with likelihood p_- , and a “moderate” value corresponding to the mean of X , with the remaining propensity $(1 - p_+ - p_-)$. Depending on the signs of $\mathbb{E}[X]$, m_+ and m_- , positive, resp. negative, these values interprets as quantification of the magnitudes of the losses, resp. of the gains (in the Insurance convention).

Notice that the (36) reduces to (35) when X is replaced by $X - \mathbb{E}[X]$. In view of the fact that

$$W_2^2(P^X, P^Y) = W_2^2(P^{X - \mathbb{E}[X]}, P^{Y - \mathbb{E}[Y]}) + (\mathbb{E}[X] - \mathbb{E}[Y])^2,$$

it would also make sense to have a discrete proxy Y which has the same mean as X .

3. A third possibility is to standardize by the mean, by simply subtracting it and then apply a two-points quantization to the centered distribution. More precisely, one could quantize $X - \mathbb{E}[X]$ to a two-point distributions, $P^Y(\cdot) = p \delta_{m_1}(\cdot) + (1 - p) \delta_{m_2}(\cdot)$, with $m_1 < m_2$. This allows to summarize the distribution of X into a mean effect $\mathbb{E}[X]$, which itself decomposes into a “moderate risk” of magnitude $m_1 + \mathbb{E}[X]$ with propensity p_1 , and a “large” risk of magnitude

$m_2 + \mathbb{E}[X]$ with propensity p_2 . This gives a quintuplet of descriptive statistics, which summarizes the characteristics of the distribution and can be thought as an alternative to Tukey’s box plot. However, compared to (36), it does not provide a propensity quantification of the mean. Such a variant may thus be deemed more useful as a descriptive statistics tool than a Risk one.

In all these variants, the rationale is to have a discrete proxy which summarizes the distribution of X and has an intuitive interpretation. The question of which variant is more sensible from a risk perspective is somehow partially a subjective issue, and is thus left to the appreciation of the reader for the application at hand. What matters is that risk comparisons between several risks be performed within the same framework.

5.3 Multivariate risk

A multivariate generalisation to risk vectors $\mathbf{X} = (X_1, \dots, X_d) \in \mathbb{R}^d$ is similar in spirit: one solves (11), with $P^{\mathbf{X}}$ a discrete two-points distribution $P^{\mathbf{X}} = (1 - p)\delta_{\mathbf{0}} + p\delta_{\mathbf{m}}$. The difference with the univariate case is that the optimal (\mathbf{m}, p) now gives one propensity measure $p_{\mathbf{X}}$ and d magnitudes measures $(m_{X_1}, \dots, m_{X_d}) =: m_{\mathbf{X}}$ to quantify the multivariate risk \mathbf{X} .

Table 1 below illustrates the approach in the case $d = 2$, for a variety of distributions. The magnitude values were obtained by direct minimization of the objective function on a sample of size $N = 5000$, for each distribution considered. The corresponding propensities were computed by counting the number of sample points in the Voronoi region.

Distribution	m_{X_1}	m_{X_2}	$p_{\mathbf{X}}$
Independence copula, Uniform marginals	0.562383	0.554198	0.8388
Minimal copula, Uniform marginals	0.50628	0.49372	1
Minimal copula, $X_1 \sim \Gamma(2, 3)$, $X_2 \sim U_{[0,1]}$	8.45238	0.296782	0.5904
Maximal copula, Uniform marginals	0.661274	0.661274	0.668
Gaussian copula $\rho = 0$, $X_1 \sim \Gamma(1.5, 2)$, $X_2 \sim \Gamma(0.5, 2)$	4.61801	1.2804	0.5348
Gaussian copula $\rho = 0.5$, $X_1 \sim \Gamma(1.5, 2)$, $X_2 \sim \Gamma(0.5, 2)$	4.54941	1.52264	0.5282
Gaussian copula $\rho = 0.95$, $X_1 \sim \Gamma(1.5, 2)$, $X_2 \sim \Gamma(0.5, 2)$	5.11324	2.03133	0.458
Gaussian copula $\rho = -0.5$, $X_1 \sim \Gamma(1.5, 2)$, $X_2 \sim \Gamma(0.5, 2)$	4.59007	0.632589	0.5384
Gaussian copula $\rho = -0.95$, $X_1 \sim \Gamma(1.5, 2)$, $X_2 \sim \Gamma(0.5, 2)$	4.92603	0.165642	0.465

Table 1: Multivariate magnitude-propensity risk. $N = 5000$.

For the first distribution with the independence copula and same uniform marginals, one gets the same magnitude measures $m_{X_1} \approx m_{X_2}$ for both marginals, as expected. Compared to the univariate case of Example 1, one gets a higher propensity and lower magnitudes. With the minimal copula, an interesting phenomenon occurs: the propensity jumps to 1, whereas magnitudes decrease to one-half. As all sample points are on the diagonal line $x_2 = 1 - x_1$, all points are closer to the mid-point $(1/2, 1/2)$ than to the origin, thus explaining the observed effect. Note that when the marginals are no longer the same, this unit propensity effect in the minimal copula case no longer occurs. With the maximal copula and same marginals, one obtains exactly the same magnitude propensity values as in the univariate case of Example 1, which is also

coherent with intuition since $X_2 = X_1$ almost surely, and the vector (X_1, X_2) thus reduces to a uniform distribution on the unit diagonal.

For a distribution with a Gaussian copula and $\Gamma(1.5, 2)$, $\Gamma(0.5, 2)$ marginals, the magnitudes reflects the difference in riskiness of the marginals. They show an increase with an increase in positive correlation. With increasing negative dependence, the magnitudes usually show some decrease, but not always, as evidenced by the increase of m_{X_1} in the last line of Table 1.

These simulations evidence a complex pattern of the influence of the dependence structure and the marginals on the magnitude-propensity measures in the multivariate case. This calls for further research.

5.4 Aggregate risk/ Portfolio optimization

In risk management, a multivariate risk $\mathbf{X} = (X_1, \dots, X_d)$ often occurs via the aggregate risk $S := \sum_{i=1}^d X_i$, or more generally via a portfolio risk $X_\beta := \beta^T \mathbf{X}$, with β in the unit simplex $\{\beta \in \mathbb{R}_+^d : \sum \beta_i = 1\}$, see e.g. McNeil, Frey, and Embrechts (2005) Chapter 6, and Rockafellar and Uryasev (2002), Uryasev and Rockafellar (2001). This reduces the study of the multivariate risk to the univariate one, so that the framework (11) can be applied without further ado.

We give an application to portfolio optimization. Let $\mathbf{X} = (X_1, X_2)$ be distributed according to a Gaussian copula with unit variances and mild correlation $\rho = 0.5$, together with Gamma marginals $X_1 \sim \Gamma(1.5, 2)$, $X_2 \sim \Gamma(0.5, 2)$. The univariate portfolio risk $X_\beta = \beta X_1 + (1-\beta)X_2$ now writes as a function of $0 \leq \beta \leq 1$. It induces a propensity-magnitude risk measure $(m_{X_\beta}, p_{X_\beta})$ for each value of β , as shown in Figure 5.4. The black labels over the circled points indicate the value of $\beta = 0, 0.05, \dots, 1$, so that $\beta = 0$ corresponds to X_β being equal to the second marginal X_2 only, while $\beta = 1$ corresponds to X_β being equal to the first marginal X_1 only. X_1 has a fatter tail and more skewed distribution than X_2 , which translates here into larger magnitude and propensity measures for $\beta = 1$ than for $\beta = 0$, see Figure 5.4.

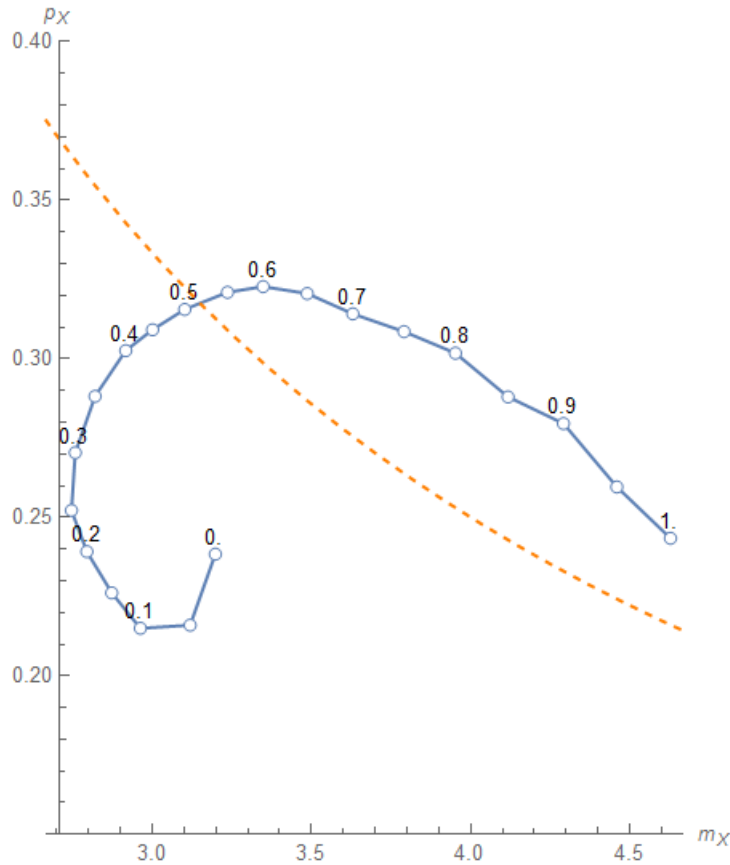


Figure 8: Magnitude-propensity plot of a portfolio $X_\beta = \beta X_1 + (1 - \beta)X_2$ (Blue circled points), with Gamma marginals, $X_1 \sim \Gamma(1.5, 2)$, $X_2 \sim \Gamma(0.5, 2)$, and a Gaussian copula for (X_1, X_2) with correlation $\rho = 0.5$, $\beta = 0, 0.05, \dots, 1$ (black labels). Expected risk curve $m_{X_\beta} p_{X_\beta} = 1$ (orange dotted curve).

The portfolio optimization problem deals with selecting the set of portfolios with acceptable/optimal risk profiles. The magnitude-propensity risk measure $(m_{X_\beta}, p_{X_\beta})$ of the portfolio can be used to formulate the portfolio optimization problem in several ways. First, it can be used to define acceptable risk zones. One can look for portfolio with a bound on the magnitude risk, the propensity risk, or a combination thereof. For example, the risk manager may decide that he is willing to accept portfolios whose expected risk, as evaluated by $m_{X_\beta} p_{X_\beta}$, is lower than some constant. The orange dotted curve on Figure 5.4 show the equation $m_{X_\beta} p_{X_\beta} = 1$. In this example, portfolios with roughly $0 \leq \beta \leq 0.5$ are deemed acceptable. Second, one can also look for the optimal/worst portfolio, in the sense that β minimizes/maximizes the expected risk $m_{X_\beta} p_{X_\beta}$. In this example, numerical inspection shows that the optimal portfolio is found for $\beta = 0.1$, and the worst one for $\beta = 0.90$. Here, it is interesting to note that the best, resp. worst, portfolio do not corresponds to $\beta = 0$, resp. $\beta = 1$. This reflects

the diversification effect.

Here, we have briefly sketched some possibilities to illustrate how the proposed framework could be applied, and leave for further research a thorough investigation of the topic.

5.5 Covariates

Another extension is to take into account the effect of covariates \mathbf{X} on a loss variable Y , by quantizing the risk of the conditional distribution $Y|\mathbf{X}$, (or a linear approximation thereof, as in quantile regression). Note that a related, but distinct, approach occurs in Frequency-Severity models (see e.g. Frees (2010) Chapter 16): there, the loss distribution has a large proportion of zeros, and the modeling is done in two parts, one for the frequency of zeros, and the other for the severity. Among other, models include the Tobit model Tobin (1958), with a censored latent variable and the individual risk model, see Frees (2010).

6 Conclusion

In this paper, we introduce the magnitude-propensity risk measures (m_X, p_X) , as a way to simultaneously quantify both the severity and the propensity of a risk X . Introduced as a particular mass transportation problem in the Wasserstein metric W_2 of the law of X to a two-points $\{0, m_X\}$ discrete distribution with mass p_X at m_X , it generalizes M -functionals, in particular the traditional risk measures like VaR and CVaR. The key observation is to view the proposed approach as a constrained optimal quantization problem, with a fixed null center. This allows to parametrize the problem with a single parameter, the optimal threshold determining the Voronoi regions, and to derive it as a solution of a fixed point equation. General existence and characterization results are obtained using B-derivatives, and sufficiency conditions for uniqueness are given. The obtained magnitude m_X has interesting interpretations in terms of classical risk measures, as (twice) a VaR or as an CVaR. In addition, it has noteworthy properties, like positive homogeneity, monotonicity w.r.t. convex order and being larger than the mean.

Visualization and comparison of risks can be done on magnitude-propensity plots, which allow for an informative comparison of risks. The effect of tails, shift in the density and outliers is reflected in the (m_X, p_X) pair. Empirically, the pair can be estimated e.g. by using a variant of Lloyd's algorithm for optimal quantization, or by direct global minimisation. This novel paradigm of visualizing and comparing risk on the bivariate magnitude-propensity scale offers a broader perspective on risk assessment and evaluation.

Eventually, the variants and extensions considered in Section 5 broadens the perspective and scope of applications of the basic approach. Multilevel quantization can help into distinguishing between “moderate” and “tail risk”, which is useful in Reinsurance. Extensions to real-valued risks allows to also quantify the “gain part” of a distribution and broadens the scope of applicability of the magnitude-propensity paradigm. Multivariate generalisation highlight some interesting phenomena, like the complex pattern of the influence of the dependence structure and the marginals on the magnitude-propensity measures in the multivariate case. A portfolio optimization problem can be formulated using the magnitude-propensity framework, and reveals an

interesting diversification effect. We hope that these variants and extensions would stimulate further research on this novel paradigm of evaluating and comparing risk.

Acknowledgments

Olivier P. Faugeras acknowledges funding from ANR under grant ANR-17-EURE-0010 (Investissements d’Avenir program).

Declaration of interest: none.

References

- Albrecher, Hansjörg, Jan Beirlant, and Jozef L Teugels (2017). *Reinsurance: actuarial and statistical aspects*. John Wiley & Sons.
- Artzner, Philippe et al. (1999). “Coherent measures of risk”. In: *Math. Finance* 9(3), pp. 203–228. ISSN: 0960-1627. DOI: 10.1111/1467-9965.00068. URL: <http://dx.doi.org/10.1111/1467-9965.00068>.
- Asmussen, Sören and Hansjörg Albrecher (2010). *Ruin probabilities*. Second. Vol. 14. Advanced Series on Statistical Science & Applied Probability. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, pp. xviii+602. DOI: 10.1142/9789814282536. URL: <https://doi.org/10.1142/9789814282536>.
- Bedford, Tim and Roger Cooke (2001). *Probabilistic Risk Analysis: Foundations and Methods*. Cambridge University Press. DOI: 10.1017/CB09780511813597.
- Bühlmann, Hans (1996). *Mathematical methods in risk theory*. Vol. 172. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Reprint of the 1970 original. Springer-Verlag, Berlin, pp. xii+210. ISBN: 3-540-61703-5.
- Embrechts, Paul et al. (2014). “An Academic Response to Basel 3.5”. In: *Risks* 2(1), pp. 25–48. ISSN: 2227-9091. DOI: 10.3390/risks2010025. URL: <http://www.mdpi.com/2227-9091/2/1/25>.
- Faugeras, Olivier P. and Ludger Rüschendorf (2018). “Risk excess measures induced by hemi-metrics”. In: *Probab. Uncertain. Quant. Risk* 3, Paper No. 6, 35. DOI: 10.1186/s41546-018-0032-0. URL: <https://doi.org/10.1186/s41546-018-0032-0>.
- Fissler, Tobias and Johanna F. Ziegel (2016). “Higher order elicibility and Osband’s principle”. In: *Ann. Statist.* 44(4), pp. 1680–1707. ISSN: 0090-5364. DOI: 10.1214/16-AOS1439. URL: <https://doi.org/10.1214/16-AOS1439>.
- Föllmer, Hans and Alexander Schied (2002). *Stochastic Finance*. Vol. 27. De Gruyter Studies in Mathematics. An introduction in discrete time. Walter de Gruyter & Co., Berlin, pp. x+422. ISBN: 3-11-017119-8. DOI: 10.1515/9783110198065. URL: <http://dx.doi.org/10.1515/9783110198065>.
- Frees, Edward W. (2010). *Regression modeling with actuarial and financial applications*. International Series on Actuarial Science. Cambridge University Press, Cambridge, pp. xviii+565. ISBN: 978-0-521-13596-2.
- Gershon, A and RM Gray (1992). “Vector quantization and signal compression Boston”. In: *Kluwer Academic*, pp. 309–315.

- Gneiting, Tilmann (2011). “Making and evaluating point forecasts”. In: *J. Amer. Statist. Assoc.* 106(494), pp. 746–762. ISSN: 0162-1459. DOI: 10.1198/jasa.2011.r10138. URL: <https://doi.org/10.1198/jasa.2011.r10138>.
- Graf, Siegfried and Harald Luschgy (2000). *Foundations of quantization for probability distributions*. Vol. 1730. Lecture Notes in Mathematics. Springer-Verlag, Berlin, pp. x+230. ISBN: 3-540-67394-6. DOI: 10.1007/BFb0103945. URL: <https://doi.org/10.1007/BFb0103945>.
- Kaas, Rob et al. (2008). *Modern actuarial risk theory: using R*. Vol. 128. Springer Science & Business Media.
- Kaplan, Stanley and B John Garrick (1981). “On the quantitative definition of risk”. In: *Risk analysis* 1(1), pp. 11–27.
- Kieffer, John C. (1983). “Uniqueness of locally optimal quantizer for log-concave density and convex error weighting function”. In: *IEEE Trans. Inform. Theory* 29(1), pp. 42–47. ISSN: 0018-9448. DOI: 10.1109/TIT.1983.1056622. URL: <https://doi.org/10.1109/TIT.1983.1056622>.
- Koenker, Roger (2005). *Quantile Regression*. Vol. 38. Econometric Society Monographs. Cambridge University Press, Cambridge, pp. xvi+349. ISBN: 978-0-521-60827-5; 0-521-60827-9. DOI: 10.1017/CB09780511754098. URL: <http://dx.doi.org/10.1017/CB09780511754098>.
- Lloyd, Stuart P. (1982). “Least squares quantization in PCM”. In: *IEEE Trans. Inform. Theory* 28(2), pp. 129–137. ISSN: 0018-9448. DOI: 10.1109/TIT.1982.1056489. URL: <https://doi.org/10.1109/TIT.1982.1056489>.
- McNeil, Alexander J., Rüdiger Frey, and Paul Embrechts (2005). *Quantitative risk management*. Princeton Series in Finance. Concepts, techniques and tools. Princeton University Press, Princeton, NJ, pp. xvi+538. ISBN: 0-691-12255-5.
- Mikosch, Thomas (2009). *Non-life insurance mathematics*. Second. Universitext. An introduction with the Poisson process. Springer-Verlag, Berlin, pp. xvi+432. ISBN: 978-3-540-88232-9. DOI: 10.1007/978-3-540-88233-6. URL: <https://doi.org/10.1007/978-3-540-88233-6>.
- Novak, Serguei Y. (2012). *Extreme value methods with applications to finance*. Vol. 122. Monographs on Statistics and Applied Probability. CRC Press, Boca Raton, FL, pp. xxvi+373. ISBN: 978-1-4398-3574-6.
- Pagès, Gilles (2018). *Numerical probability*. Universitext. An introduction with applications to finance. Springer, Cham, pp. xxi+579. DOI: 10.1007/978-3-319-90276-0. URL: <https://doi.org/10.1007/978-3-319-90276-0>.
- Pollard, David (1982). “Quantization and the method of k -means”. In: *IEEE Trans. Inform. Theory* 28(2), pp. 199–205. ISSN: 0018-9448. DOI: 10.1109/TIT.1982.1056481. URL: <https://doi.org/10.1109/TIT.1982.1056481>.
- Puccetti, Giovanni (2013). “Sharp bounds on the expected shortfall for a sum of dependent random variables”. In: *Statist. Probab. Lett.* 83(4), pp. 1227–1232. ISSN: 0167-7152. DOI: 10.1016/j.spl.2013.01.022. URL: <https://doi.org/10.1016/j.spl.2013.01.022>.
- Rachev, Svetlozar T and Ludger Rüschendorf (1998). *Mass transportation problems. Vol. I*. Vol. 1. Probability and its Applications (New York). Theory. Springer-Verlag, New York, pp. xxvi+508. ISBN: 0-387-98350-3.

- Robinson, Stephen M. (1987). “Local structure of feasible sets in nonlinear programming. III. Stability and sensitivity”. In: 30. Nonlinear analysis and optimization (Louvain-la-Neuve, 1983), pp. 45–66. DOI: 10.1007/bfb0121154. URL: <https://doi.org/10.1007/bfb0121154>.
- Rockafellar, R. Tyrrell and Stanislav Uryasev (2002). “Conditional value-at-risk for general loss distributions”. In: *Journal of banking & finance* 26(7), pp. 1443–1471.
- Rootzén, H. and C. Klüppelberg (1999). “A single number can’t hedge against economic catastrophes”. In: *Ambio* 28, pp. 550–555.
- Rüschendorf, Ludger (2013). *Mathematical Risk Analysis*. Springer Series in Operations Research and Financial Engineering. Dependence, risk bounds, optimal allocations and portfolios. Springer, Heidelberg, pp. xii+408. DOI: 10.1007/978-3-642-33590-7. URL: <http://dx.doi.org/10.1007/978-3-642-33590-7>.
- Santambrogio, Filippo (2015). “Optimal transport for applied mathematicians”. In: *Birkhäuser, NY*, pp. 99–102.
- Scholtes, Stefan (2012). *Introduction to piecewise differentiable equations*. Springer-Briefs in Optimization. Springer, New York, pp. x+133. ISBN: 978-1-4614-4339-1. DOI: 10.1007/978-1-4614-4340-7. URL: <https://doi.org/10.1007/978-1-4614-4340-7>.
- Shorack, Galen R. (2000). *Probability for statisticians*. Springer Texts in Statistics. Springer-Verlag, New York, pp. xviii+585. ISBN: 0-387-98953-6.
- Tobin, James (1958). “Estimation of relationships for limited dependent variables”. In: *Econometrica: journal of the Econometric Society*, pp. 24–36.
- Trushkin, Alexander V. (1982). “Sufficient conditions for uniqueness of a locally optimal quantizer for a class of convex error weighting functions”. In: *IEEE Trans. Inform. Theory* 28(2), pp. 187–198. ISSN: 0018-9448. DOI: 10.1109/TIT.1982.1056480. URL: <https://doi.org/10.1109/TIT.1982.1056480>.
- Uryasev, Stanislav and R. Tyrrell Rockafellar (2001). “Conditional value-at-risk: optimization approach”. In: *Stochastic optimization: algorithms and applications (Gainesville, FL, 2000)*. Vol. 54. Appl. Optim. Kluwer Acad. Publ., Dordrecht, pp. 411–435. DOI: 10.1007/978-1-4757-6594-6_17. URL: https://doi.org/10.1007/978-1-4757-6594-6_17.
- Villani, Cédric (2003). *Topics in Optimal Transportation*. Vol. 58. Graduate Studies in Mathematics. American Mathematical Society, pp. xvi+370. ISBN: 0-8218-3312-X. DOI: 10.1007/b12016. URL: <http://dx.doi.org/10.1007/b12016>.
- Villani, Cédric (2009). *Optimal transport*. Vol. 338. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Old and new. Springer-Verlag, Berlin, pp. xxii+973. ISBN: 978-3-540-71049-3. DOI: 10.1007/978-3-540-71050-9. URL: <http://dx.doi.org/10.1007/978-3-540-71050-9>.
- Wang, Shaun (1996). “Premium Calculation by Transforming the Layer Premium Density”. In: *ASTIN Bulletin* 26(1), 71–92. DOI: 10.2143/AST.26.1.563234.

Wolfram Research, Inc. (2020). *Mathematica*. Version 12.1. URL: <https://www.wolfram.com>.