Risk Quantization by Magnitude and Propensity

Olivier P. Faugeras, Gilles Pages

Risk Quantization by Magnitude and Propensity

Olivier P. Faugeras*[1] and Gilles Pagès[2]

[1]Toulouse School of Economics, Université Toulouse 1 Capitole, 1, Esplanade de l'Université. Bureau T106. 31080 Toulouse Cedex 06, France.
olivier.faugeras@tse-fr.eu

[2]Laboratoire 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@upmc.fr

May 22, 2021

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 hign 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 Expected shortfall, which typically conflate both effects. In its simplest form, (m_X, p_X) is obtained by mass transportation in Wasserstein metric of the law P^X 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 proposed approach.

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 risk[1] X by a risk measure $\rho(X)$, which can be thought of as a one-sided deterministic univariate summary of

*Corresponding author

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

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 $\delta_{\rho(X)}(.)$ 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 measures 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. 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, we give a conclusion of the main results in Section 5 and add some perspective for further research about possible variants and extensions of the basic approach.

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).

Risk due to high magnitudes and risk due to high propensities manifests itself on several levels: intrinsically for a single risk variable, relatively in the comparison of two risks, and in the stochastic order itself. The best way to illustrate our point is to give examples stripped to their simplest form, i.e. for two-points discrete measures. Hence, we have the following intentionally simplistic examples:

- **Intrinsic magnitude-propensity aspect for a single random risk:** Let X_1 s.t. $P(X_1 = 1000) = 0.1$, $P(X_1 = 0) = 0.9$, while X_2 is s.t. $P(X_2 = 200) = 0.5$, $P(X_2 = 0) = 0.5$. X_1 and X_2 have the same mean $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 100$. X_1 has a "high" magnitude risk 1000 of "low" propensity 0.1 while X_2 has a "small" magnitude risk 200 of "high" propensity 0.5, see Figure 1.

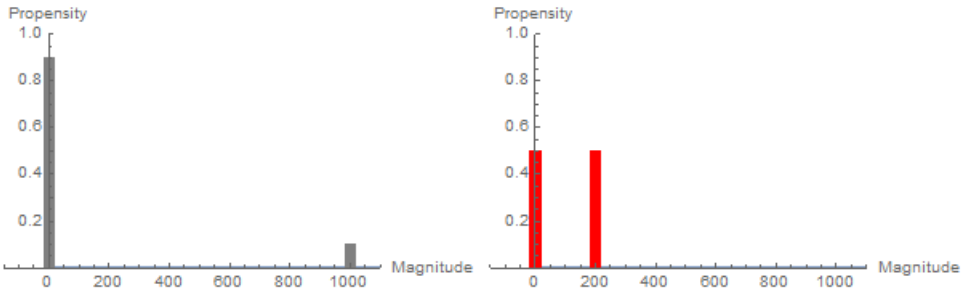

Figure 1: The dual nature of risk: X_1 has high magnitude risk (left) vs X_2 has high propensity risk (right). Yet, $\mathbb{E}[X_1] = \mathbb{E}[X_2]$.

- **Relative magnitude-propensity effect for the comparison of risks:**
  When one considers a pair (X_1, X_2) of risk variables and regards risk as a relative property of one r.v. X_1 w.r.t. the other r.v. X_2, the magnitude-propensity duality manifests itself in the comparison of risks and in the ordering structure of probability measures. This is illustrated in the following (also intentionally simplistic) examples:

  Example 1. *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 due to a difference in magnitudes, see Figure 2.*

  Example 2. *Let X_1 s.t. $P(X_1 = 0) = 0.95$, $P(X_1 = 1000) = 0.05$, while X_2 is s.t. $P(X_2 = 0) = 0.5$, $P(X_2 = 1000) = 0.5$, see Figure 3. Then, X_1 and X_2 have same magnitudes, but X_2 is more risky than X_1 due to a difference in propensities, see Figure 3.*

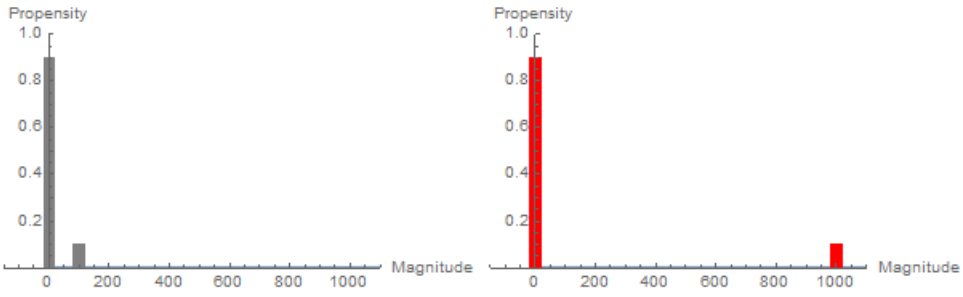

Figure 2: Comparison of risks for Example 1: X_2 (right) is riskier than X_1 due to a difference of magnitudes

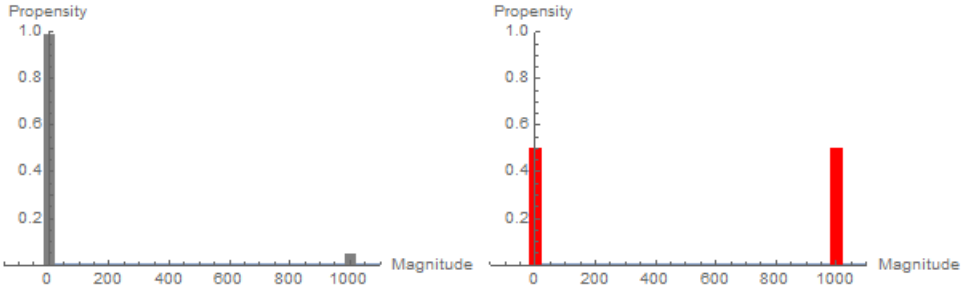

Figure 3: Comparison of risks for Example 2: X_2 (right) is riskier than X_1 due to a difference of propensities

- **Magnitude-propensity effect in the stochastic order:**
  In both cases of Examples 1 and 2, one has

$$F_{X_1}(x) \geq F_{X_2}(x), \quad \forall x \in \mathbb{R}^+, \tag{1}$$

  so that $X_1 \prec_{st} X_2$, where $\prec_{st}$ stands for the usual stochastic order. Yet, (1) does not distinguish between a horizontal shift of c.d.f.s due to a difference in magnitudes (Figure 4 left), and a vertical shift of c.d.f.s due to a difference of propensities (Figure 4 right).

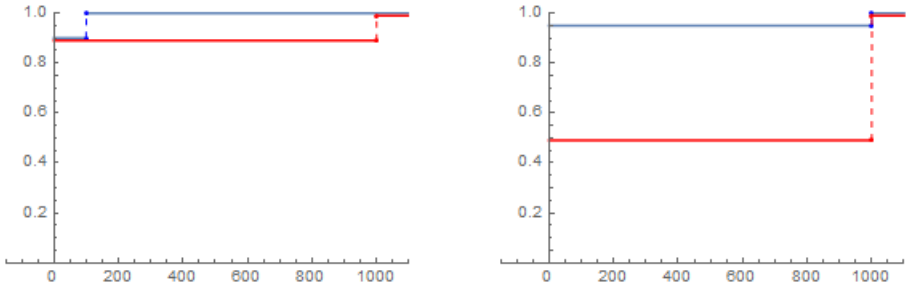

Figure 4: C.d.f.s of X_1 (blue), X_2 (red). Left: Example 1, difference of magnitudes (horizontal shift). Right: Example 2, differences of propensities (vertical shift).

Remark 1. *From the mathematical standpoint, let us remark that this magnitude-propensity duality of risk can be formalized mathematically by the concept of a Galois connection between the magnitude and probabilities spaces, considered as two ordered spaces,* $(\mathbb{R}, \leq)$ *and* $([0,1], \leq)$*. Indeed, denote by* F_X*, resp.* Q_X*, the cumulative distribution function, resp. the quantile function, of* X*. Then,* (F_X, Q_X) *forms a Galois connection between the magnitudes space* $(\mathbb{R}, \leq)$ *and the propensity space* $([0,1], \leq)$*, i.e. for all* $x \in \mathbb{R}$ *and* $t \in (0,1)$*,*

$$t \leq F_X(x) \Leftrightarrow Q_X(t) \leq x.$$

See O. P. Faugeras and Rüschendorf (2017) for details and a generalization of the concept of quantile to the multivariate case.

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

$$\rho : \mathscr{X} \to \mathbb{R}^+, \tag{2}$$

where $\mathscr{X}$ 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. Rüschendorf (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 Hansjörg 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 Olivier P. Faugeras and Rüschendorf (2018)). These numerous approaches have lead 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,

$$VaR_\alpha(X) := \inf\{x : P(X \leq x) \geq \alpha\}, \quad 0 < \alpha < 1 \tag{3}$$

  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,

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

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

- The Expected Shortfall, (also called the Conditional Value at Risk, see Uryasev and R. Tyrrell Rockafellar (2001)),

$$ES_\alpha(X) := \mathbb{E}[X|X \geq VaR_\alpha(X)], \tag{4}$$

  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 2 (On elicitability)**.** *It has been argued in the literature that elicitability is also a desirable property for risk measures (see Embrechts et al. (2014)). Elicitability 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 expected shortfall (ES) is a coherent law invariant risk measure, but is not elicitable. ES is jointly elicitable with VaR, see Fissler and Ziegel (2016).*

This consideration of elicitability suggests the use of bivariate risk measure (ES,VaR), i.e. of combining magnitude and propensity type univariate risk measures to obtain something that is both coherent and elicitable. This is another supplementary motivation for arguing that one should switch to bivariate risk measures for the study of univariate risks.

Remark 3 (On parametrized risk measures)**.** *For risk measures depending on a parameter, like α in $VaR_\alpha(X)$ or $ES_\alpha(X)$, 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 ES give in fact curves, $\alpha \to VaR_\alpha(X)$ and $\alpha \to ES_\alpha(X)$. 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 $\overline{F}_X$, 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:

$$\mathcal{T}_c(P^X, P^Y) := \inf_{P^{(X,Y)}\in\mathcal{P}(P^X,P^Y)} \int c(x,y) P^{(X,Y)}(dx, dy), \tag{5}$$

where $c : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$ 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), Cédric Villani (2003), Cédric 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 $\mathcal{P}(P^X, \delta_m) = \{P^X(dx) \times \delta_m(dy)\}$. (5) thus simplifies as the expected cost between $X \sim P^X$ and a fixed point m,

$$\mathcal{T}_c(P^X, \delta_m) = \int c(x,m)P^X(dx) = \mathbb{E}\, c(X,m). \tag{6}$$

Therefore, minimizing the transportation cost (5) over the set $\mathcal{D} := \{\delta_m(dy), m \in \mathbb{R}\}$ of Dirac measures is equivalent to minimizing the expected cost (6) over $m \in \mathbb{R}$,

$$\mathcal{T}_c(P^X, \mathcal{D}) = \inf_{P^Y \in \mathcal{D}} \mathcal{T}_c(P^X, P^Y) = \inf_{m \in \mathbb{R}} \mathbb{E}\, c(X,m). \tag{7}$$

In particular,

- for $c(x,y) = (x-y)^2$, (5) is the squared L_2-Wasserstein metric W_2 and (7) writes as the variance,

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

  and is obviously minimised for the mean $m = EX$. In addition, when P^X is replaced by the conditional law of X given $X \geq VaR_\alpha(X)$, one obtains the Expected shortfall (4).
- 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 and Basset), that is to say the Value-At-Risk (3).
- For $c(x,y) = y + \frac{(x-y)\mathbb{1}_{x \geq y}}{1-\alpha}$, one gets simultaneously the Expected Shortfall and the Value-at-risk, when $\mathbb{E}[X] < \infty$: the optimal value in (7) is the Expected Shortfall while the argmin gives the Value-At-Risk, see Uryasev and R. Tyrrell Rockafellar (2001), R Tyrrell Rockafellar and Uryasev (2002).
- And so on for other statistical functionals. See also Olivier P. Faugeras and Rüschendorf (2018) for optimal transportation induced by cost functions which are hemi-metrics encoding an order.

The above discussion shows that the statistical functionals and risk measures, which can be expressed as an M-estimator solving (7) 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 (7), but with the target Dirac distribution δ_m replaced by a two-points distribution P^Y,

$$P^Y = (1-p)\delta_0 + p\delta_m. \tag{8}$$

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 (8). 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 $\mathcal{A}_0$,*

$$(m_X, p_X) = \arg \inf_{P^Y \in \mathcal{A}_0} W_2(P^X, P^Y). \tag{9}$$

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 at $m > 0$ 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 (9) 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,

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

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

$$W_2^2(P^X, P^Y) = \int_0^1 (Q_X(t) - Q_Y(t))^2 \, dt, \tag{10}$$

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 (10) 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 (9), then it satisfies the following system of equations,

$$m_X = 2Q(1 - p_X), \tag{11}$$

$$m_X = \frac{\int_{1-p_X}^1 Q(t)dt}{p_X}. \tag{12}$$

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

$$\frac{2p_X}{f(Q(1 - p_X))} - Q(1 - p_X) > 0. \tag{13}$$

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 (9) 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 (11) and (12).

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

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

With (m_X, p_X) satisfying (11), (12), the condition writes

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

which is (13).

□

Remark 4. *1. If $1 - p_X$ is in the range of F (in particular if F is continuous), then, (11) writes*

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

and by Exercise 3.3 in Shorack (2000) p. 113, (12) writes as

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

This equation will be derived in the general case from the optimal quantization viewpoint, see Theorem 3.5 and Remark 5 below.

2. Setting $a := Q(1-p_X)$, the sufficiency condition (13) writes also

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

In view of the fact that $\mathbb{E}[X] = \int_0^\infty t f(t) = \int_0^\infty (1 - F(t))dt < \infty$, 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 3 (Uniform distribution)**.** *For $X \sim U_{[0,a]}$, i.e. $Q(p) = ap$, (11) and (12) give $p_X = 2/3$ and $m_X = 2a/3$.*

Example 4 (Exponential distribution)**.** *For $X \sim Exp(\lambda)$, with $\lambda > 0$, the memory-lessness 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 does is fixed and does not depend on the parameter (a, resp. λ) of the distribution.

Example 5 (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 (22), one finds that the optimal threshold is $a = \frac{1}{\theta - 2}$. Hence, (23) 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 (9) 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 $(\mathbb{R}^d, ||.||)$ is a mapping $T : \mathbb{R}^d \mapsto \{x_1, \ldots, x_N\}$, where $\{x_1, \ldots, x_N\}$ 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

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

and is determined by the pairs $\{(x_i, A_i), i = 1, \ldots n\}$. 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

$$D(T; P^X) = D_N(x_1, \ldots, x_N) := \mathbb{E} \min_{x_i} ||X - x_i||^2 \tag{14}$$

If the support of P^X has at least N elements, existence of an N-optimal quantizer follows from the fact that $(x_1, \ldots, x_N) \to \sqrt{D_N(x_1, \ldots, x_N)}$ is 1-Lipshitz and a non-trivial compacity 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 $x^N := (x_1, \ldots, x_N)$ 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

$$\frac{\partial D_N}{\partial x_i}(x^N) = \mathbb{E}[2(x_i - X)\mathbb{1}_{X \in A_i}]. \tag{15}$$

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 $T : \mathbb{R}^+ \mapsto \{0, m\}$ with $T(x) = m\mathbb{1}_{x \geq a}$, where a is a threshold to determinate. Then, the optimal quantization problem with constrained knot at zero writes

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

By the results of the previous section (see (14)), 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 (14) the distortion/objective function writes as a sole function of the magnitude m, as

$$L(m) := \mathbb{E}[X^2 \wedge (X - m)^2]. \tag{16}$$

We first study the differentiability properties of the objective function (16) 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 $H : x \to X^2 \wedge (X-x)^2$ 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 $f : \mathbb{R} \to \mathbb{R}$ is Bouligand differentiable (B-differentiable) at x_0 if there exists a positive homogeneous function $\nabla^B f(x_0) : \mathbb{R} \to \mathbb{R}$ s.t.*

$$f(x_0 + v) = f(x_0) + \nabla^B f(x_0)(v) + o(v), \quad \forall v \in \mathbb{R}. \tag{17}$$

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 $h_1, h_2 : \mathbb{R} \to \mathbb{R}$ 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. $\forall |v| < \delta$,

$$h_1(x + v) \leq h_1(x) + \varepsilon_1 \tag{18}$$
$$h_2(x + v) \geq h_2(x) - \varepsilon_2 \tag{19}$$

$h_1(x) + \varepsilon_1 = h_2(x) + \varepsilon_1 - d$, therefore (18) and (19) 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

$$\begin{aligned} h_1(x + v) &= h_1(x) + h_1'(x)v + o(v) \\ h_2(x + v) &= h_2(x) + h_2'(x)v + o(v) \end{aligned} \tag{20}$$

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.

□

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

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

and, for $x > 0$,

$$\nabla^B L(x)(v) = 2v\,\mathbb{E}[(x-X)\mathbb{1}_{X>x/2}] + P(X = x/2)\min(0, xv). \tag{21}$$

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 (20) 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 (17) 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,

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

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

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

and for $x > 0$,

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

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

□

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 (9).*

ii) An optimal magnitude-propensity pair (p_X, m_X) is characterized as a solution of the equation

$$2a = \mathbb{E}[X|X > a], \quad a > 0, \tag{22}$$

and then

$$m_X = 2a, \quad p_X = P(X > m/2). \tag{23}$$

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

Proof. i) By the results of the Section 3.2, the objective function L of (16) is locally Lipshitz 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 $\nabla^B L(.)(.)$ 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

$$\mathbb{E}[(m_X - X)\mathbb{1}_{X > m_X/2}] = 0. \tag{24}$$

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

$$2\,\mathbb{E}[(m_X - X)\mathbb{1}_{X > m_X/2}] + m_X P(X = m_X/2) = 0. \tag{25}$$

Plugging (24) into (25) implies that $P(X = m_X/2) = 0$ and yields (22), upon reparametrizing by $a = m_X/2$.

□

Remark 5.

i) We hopefully obtain the same formula between (22) of Theorem 3.5 and (11), (12) 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 (21), 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

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

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 $x \to x^3 f(x)$, 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, $0 \in \textit{supp}(P^X)$ and $\mathbb{E}[X^2] < \infty$. Assume*

i) $\tilde{f} : x \to x^3 f(x)$ is strictly log-concave on $\mathbb{R}_+$;

ii) $\lim_{x\to 0} xf(x) = 0$.

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 $\tilde{f} : x \to x^3 f(x)$ has a limit in $[0, \infty)$ as $x \to 0$. Thus, we may set $\tilde{f}(0) = \lim_{x\to 0} x^3 f(x) = 0$, 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 $0 \in \mathtt{supp}(P^X)$, 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

$$\Phi(y) = 2y\overline{F}(y) - \overline{K}(y), \tag{26}$$

where we have set $\overline{F}(y) = \int_y^b f(t)dt$ and $\overline{K}(y) = \int_y^b tf(t)dt$. 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

$$\Phi'(y) = 2\overline{F}(y) - yf(y). \tag{27}$$

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

$$\begin{aligned}\Phi''_r(y) &:= -3f(y) - yf'_r(y) = -yf(y)\left(\frac{3}{y} + \frac{f'_r}{f}(y)\right)\\ &= -yf(y)\zeta(y),\end{aligned} \tag{28}$$

where

$$\zeta(y) := 3/y + f'_r(y)/f(y) \tag{29}$$

with f'_r denoting the right-handed derivative of f. By assumption, $x \to x^3 f(x)$ 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 $\lim_{y\to b} \zeta(y) = \ell \geq 0$. 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), $\Phi'(0) = 2\overline{F}(0) - \lim_{y\to 0} yf(y) = 2 > 0$.

Case ia) if $b = \infty$, then $\lim_{y\to\infty} \Phi'(y) = -\lim_{y\to\infty} yf(y) = 0$ since $\mathbb{E}[X] < \infty$. Thus $\Phi' > 0$ and Φ is increasing on $(0, \infty)$. But since $\Phi(0) = -\mathbb{E}[X] < 0$ and $\lim_{y\to\infty} \Phi(y) = 0$, 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 $\lim_{y\to b} \Phi'(y) = -bf(b-) \leq 0$. 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, $f(b-) > 0$, $\lim_{y\to b} \Phi'(y) < 0$ and there exists a unique $y_1 \in (0, b)$ s.t. $\Phi'(y_1) = 0$.

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

Case iia) If $\lim_{y\to 0}\zeta(y)=\eta\leq 0$. 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 $\lim_{y\to b}\Phi'(y)=0$, resp. $=-bf(b-)\leq 0$, for $b=\infty$, resp. $b<\infty$. Thus, necessarily $\lim_{y\to 0}\zeta(y)=\eta>0$.

Case iib) If $\lim_{y\to 0}\zeta(y)=\eta>0$. 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, $\lim_{y\to\infty}\Phi(y)=0$ 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 $\lim_{y\to\infty}\Phi'(y)=0$ 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 $\lim_{y\to b}\Phi(y)=0$, 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 $(0,y_1)$. □

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

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

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

□

Remark 6. *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 7. *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,*

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

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

$$\mathbb{E}[(m_X - X)\mathbb{1}_{X \geq m_X/2}] = 0. \tag{30}$$

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 (30) by the change of variable $x = -t$ and the symmetry of $\tilde{f}$. For $x_3 = m_X$, the stationary condition is also (30) 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 (23) or (11), (12) has several interesting characteristics. From (11), resp. (12), it is clear that the magnitude m_X can be interpreted either as (twice) a Value-At-Risk $Var_\alpha(X)$, resp. as an Expected Shortfall $ES_\alpha(X)$, for a special value of α:

$$m_X = ES_{p_X}(X) = 2VaR_{1-p_X}(X). \tag{31}$$

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 3 about the "right" choice of α for parametrized risk measures like $VaR_\alpha(X)$ or $ES_\alpha(X)$: 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.

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 (14) 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. Rüschendorf (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 (22), is lower than the corresponding optimal threshold a_Y for Y. Therefore, $m_X \leq m_Y$. □

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

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

as $\overline{K}(x) \geq x\overline{F}(x)$. 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 $m_X \geq \mathbb{E}[X]$. □

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 the magnitude and propensity axes. Figure 5 illustrates the resulting magnitude-propensity plot, for the uniform, exponential and Pareto distributions of Examples 3-5. 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 3-5. However, one can use the computation capabilities of, e.g., Mathematica Wolfram Research, Inc. (2020) to obtain a closed-form

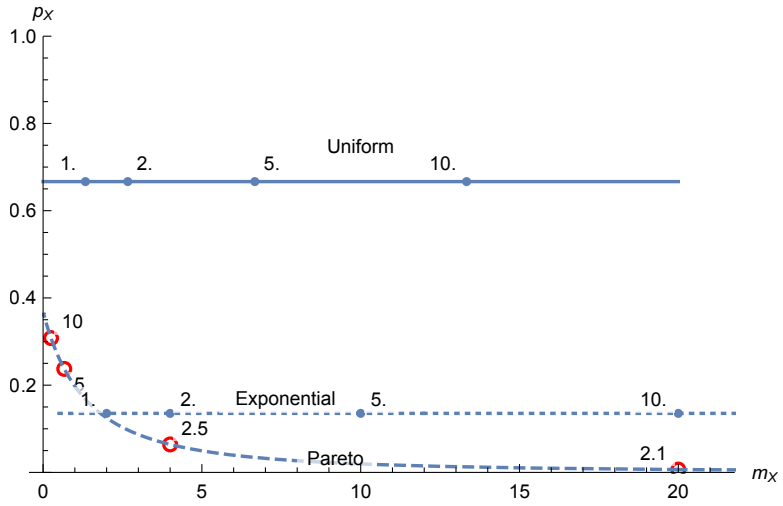

Figure 5: Magnitude-propensity plots for the uniform $U_{[0,a]}$ distribution (solid line), Exponential $Exp(\lambda)$ (dotted) and Pareto $Pa(\theta)$ (dashed), for varying values of the parameter a, λ, θ.

expression of the $\mathbb{E}[X|X > a]$ and then find numerically the root of (22). We illustrate the procedure for the Gamma and Weibull distributions, two distributions frequently encountered in insurance mathematics.

Example 6 (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 (22) numerically with initial point $\mathbb{E}[X]$. The resulting magnitude-propensity plot is displayed in Figure 6.

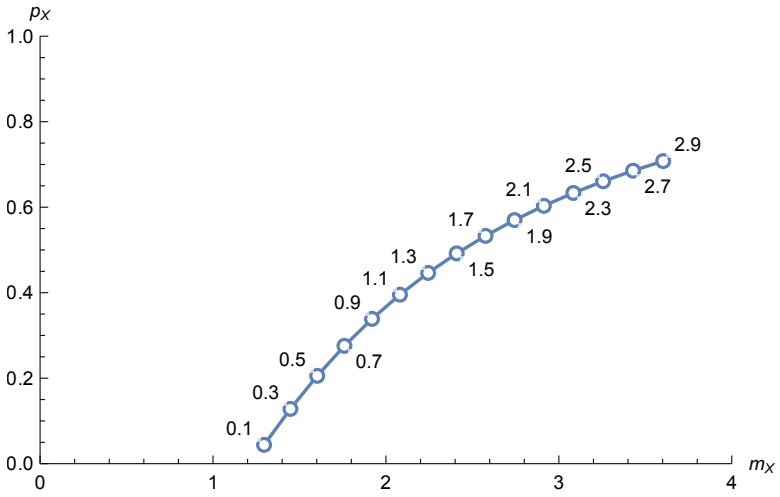

Figure 6: Magnitude-propensity plot for Gamma distribution $X \sim \Gamma(\alpha, 2)$, for α varying from 0.1 to 2.9. The corresponding mean $\mathbb{E}[X]$ is indicated near the circled (m_X, p_X) points.

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 5. 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 7. Hence, both the magnitude and propensity coding the risk borne by X will increase.

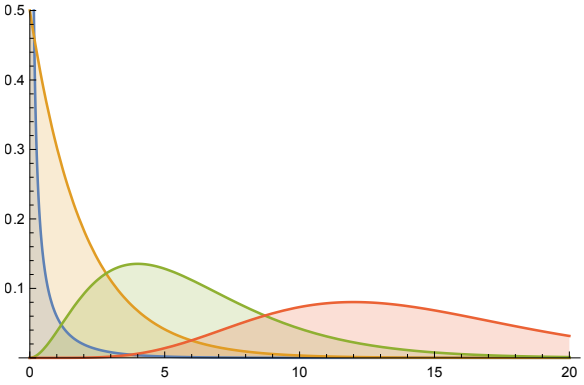

Figure 7: Probability density functions f for Gamma distributions $\Gamma(\alpha, 2)$, with $\alpha = 0.1$ (blue), 1 (orange), 3 (green), 7 (red).

Example 7 (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 8 below.*

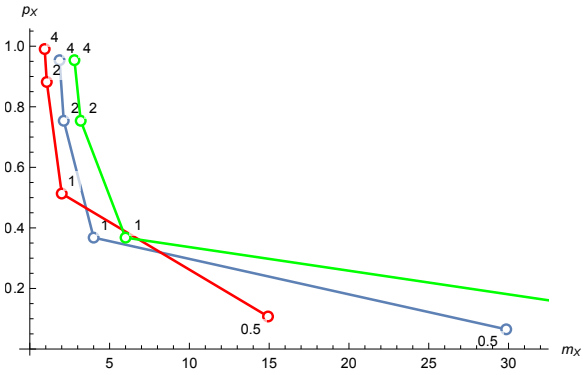

Figure 8: Magnitude-propensity plot for a Weibull $W(\alpha, \beta)$ distribution for $\beta = 2$ (blue), $\beta = 1.5$ (red), $\beta = 3$ (green). The magnitude-propensity (m_X, p_X) points are labeled by the values of $\alpha = 0.5$, 1, 2, 4.

4.3 Empirical computations and illustrations

The characterization (22) shows that the optimal threshold is a fixed point of the function

$$a \mapsto \frac{\mathbb{E}[X|X \geq a]}{2}. \tag{32}$$

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 (32) 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 Appendix 5.2. 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 (16), 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 3, we known 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 for 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.

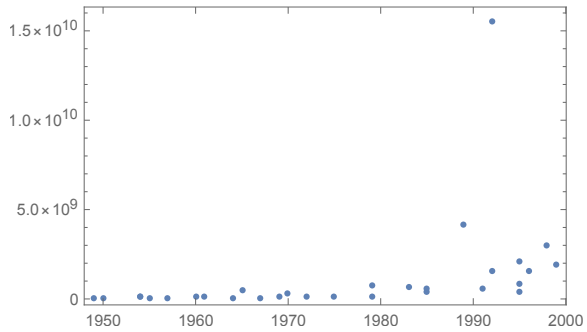

Figure 9: US Hurricane Losses data (Reported losses) in the U.S. $1949--1999$. Available at https://datarepository.wolframcloud.com/resources/US-Hurricane-Loss.

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 Conclusion

5.1 Summary

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 Expected Shortfall. 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 Expected Shortfall. 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.

5.2 Perspective: towards general risk quantization

To make the presentation clear, we focus in this paper on the simplest way to quantify risk on the magnitude and propensity scale, i.e. on the basic idea (9) with a constrained two-points quantizer $\{0, m_X\}$. This view of measuring risk via a constrained optimal quantization problem 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 these possible variants and extensions.

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.

(a) Variant: Mean standardization / Moderate-Large risks.

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

One could also standardize differently by the mean, by simply subtracting it. More precisely, one could quantize $X - \mathbb{E}[X]$ to a two-point distributions, $P^Y(.) = p\delta_{m_1}(.) + (1-p)\delta_{m_2}(.)$, 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 $m_1 + \mathbb{E}[X]$ and propensity p_1, and a "large" risk of magnitude $m_2 + \mathbb{E}[X]$ and 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.

(b) Variant: three-points quantification.

Also, one can refine our proxy by looking for a quantization to a discrete distribution with more than two points. For example, another way to refine one's measure of risk into a "moderate" risk and a "large" risk (or, say, "Low" risk and "Tail" risk), is to use directly 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 both moderate, resp. large risk, in the magnitude and propensity scale with (m_1, p_1), resp. (m_2, p_2).

Such a simultaneous quantization of the "mild" and "extreme" risk parts of an insurance risk X could reveal to be very valuable in the field of Reinsurance, see

e.g. Hansjörg 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 risk quantization approach by magnitude and propensity seems to be particularly well-suited for this task.

(c) Extension to financial risk.

For X real-valued (i.e. we take the financial mathematics convention, with $X \geq 0$ standing for a gain and $X < 0$ for a loss), one can consider for example a three-points risk quantization with the following distribution,

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

with $m_- \leq \mathbb{E}[X] \leq m_+$ or with

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

with the constraint $\mathbb{E}[X] = \mathbb{E}[Y]$, $m_- \leq 0 \leq m_+$. This allows to summarize the gain/loss of the distribution on both the magnitude and propensity scales.

(d) Multivariate risk.

Multivariate generalisations to risk vectors $\mathbf{X} \in \mathbb{R}^d$ are similar in spirit, although the computations are less explicit. One can somehow reduce to the scalar case by considering the risk quantization of a portfolio vector $\boldsymbol{\beta}^T \mathbf{X}$, for $\boldsymbol{\beta}$ in unit simplex, see e.g. R Tyrrell Rockafellar and Uryasev (2002), Uryasev and R. Tyrrell Rockafellar (2001).

(e) 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).

Acknowledgments

Olivier P. Faugeras acknowledges funding from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program).
Declaration of interest: none.

Appendix: Mathematica code

Empirical computation of (m_X, p_X) by Lloyd's Method:

```
(*Input: data= dataset;
        x0=starting value of the threshold search (take the mean
        of the data set for  example)
```

```
  Ouput: {m_X,p_X}
*)
  mplloyd[data_, x0_] :=
  Module[{a},
      iteratefunction[a_] := Mean[Select[data, # > a &]]/2;
      a = FixedPoint[iteratefunction[#] &, x0];
      {2 a, NProbability[x > a, x \[Distributed]
      EmpiricalDistribution[data]]}
        ]
  (* one could also have computed p_x by
  1-CDF[EmpiricalDistribution[data],a] *)
```

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/CBO9780511813597.

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, O. P. and Ludger Rüschendorf (2017). "Markov morphisms: a combined copula and mass transportation approach to multivariate quantiles". In: *Mathematica Applicanda*. 1st ser. 45, pp. 3–45.

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 elicitability 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.

Gersho, 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.

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.

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.

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: *Birkäuser, NY*, pp. 99–102.

Scholtes, Stefan (2012). *Introduction to piecewise differentiable equations*. SpringerBriefs 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.