Distortion risk measures, ROC curves, and distortion divergence

Schumacher, J.M.

DOI
10.2139/ssrn.2956334

Publication date
2017

Document Version
Submitted manuscript

Distortion Risk Measures, ROC Curves, and Distortion Divergence

Johannes M. Schumacher*

October 18, 2017

Abstract

Distortion functions are employed to define measures of risk. Receiver operating characteristic (ROC) curves are used to describe the performance of parametrized test families in testing a simple null hypothesis against a simple alternative. This paper provides a connection between distortion functions on the one hand, and ROC curves on the other. This leads to a new interpretation of some well known classes of distortion risk measures, and to a new notion of divergence between probability measures.

1 Introduction

The theory of risk measures is rooted in the literature on actuarial premium principles (see for instance Goovaerts et al. (1984, Ch. 3)), and has been studied intensively since the seminal work of Artzner et al. (1999). Generally speaking, risk measures (also known as insurance premium principles, or as monetary utility functions) are mappings that assign a real number to a given random variable. The term "coherent risk measure" as introduced by Artzner et al. (1999) refers to a set of axioms that such a mapping may satisfy. The coherence axioms include monotonicity, subadditivity, additive homogeneity with respect to constants ("translation invariance"), and multiplicative homogeneity with respect to the positive reals. Adding the further axioms of determination by the distribution ("distribution invariance") and comonotonic additivity, one obtains the class of so called *distortion risk measures*. Subject to a continuity assumption, these risk measures can be characterized as the ones that can be written in the form $\rho(X) = \inf_{Q \in \mathcal{Q}} E^Q[X]$, where the collection of probability measures $\mathcal{Q}$ is of the following type: for some given probability measure P and concave distortion function $g : [0,1] \to [0,1]$, the set $\mathcal{Q}$ consists of all measures that satisfy $Q(A) \leq g(P(A))$ for all events A. Alternatively, distortion risk measures can be described as integrals with respect to non-additive measure. For full details, the reader may refer to Choquet (1954); Schmeidler (1986); Denneberg (1994); Wang et al. (1997); Denuit et al. (2005); Föllmer and Schied (2011).

*University of Amsterdam, Faculty of Economics and Business, Section Quantitative Economics; `j.m.schumacher@uva.nl`. The author gratefully acknowledges comments by two anonymous reviewers, which helped to improve the paper.

Distortion risk measures have been used extensively in various contexts; see for instance Wirch and Hardy (1999); Young (1999); Tsanakas (2004); Hürlimann (2004); Sereda et al. (2010); Assa (2015); Boonen (2015). They have been related to Yaari's dual theory of choice under risk (Wang, 1996) and to the theory of ordered weighted averaging operators (Belles-Sampera et al., 2013). Many proposals for specific distortion functions have been made; see for instance Wang (1995, 1996, 2000); Reesor and McLeish (2003); Cherny and Madan (2009); Hürlimann (2014). These proposals in fact specify *families* of distortion functions, in which the amount of allowed distortion is controlled by one or more parameters. A well known example is the CVaR family (conditional value-at-risk, also known as expected shortfall), which is used in the Basel III regulations on minimum capital requirements for banks (Bank for International Settlements, 2016).

The purpose of this paper is to discuss the connection between distortion risk measures on the one hand and binary hypothesis testing problems on the other hand. Intuitively, distortion functions are used to describe a collection of probability measures that are "not too far" from a given reference measure. From the point of view of hypothesis testing, two probability measures are farther apart when it becomes easier to design good tests that distinguish between them, or in other words, when the tradeoff between type-I and type-II error probabilities is relatively nonrestrictive. For an assessment of this tradeoff, one may compute, for each given significance level, the maximally achievable power of a test of the given null against the given alternative. The curve that is obtained by plotting the relation between significance level and maximal power is known as the optimal *operating characteristic curve* or *receiver operating characteristic curve* (ROC curve); see for example Dudewicz and Mishra (1988, p. 441). Like distortion functions used for risk measures, optimal ROC curves are nondecreasing and concave functions that map the unit interval onto itself. By employing the optimal ROC curve for a particular testing problem as a distortion function, one obtains a distortion risk measure. The risk measure that is obtained in this way may be said to have been calibrated on the given testing problem. This specific origin does not impair its general applicability. As is shown below, standard testing problems involving normal, exponential, uniform and Bernoulli distributions are related by this calibration to popular distortion risk measures, including CVaR.

While ROC curves can therefore be used to generate distortion functions, the connection between ROC curves and distortion risk measures goes deeper than that. It is shown in Prop. 3.1 below that the set of "test measures" associated with a given distortion function can be described directly in terms of the ROC curve. The latter description is convenient, because it is based on functions that are defined on the interval $[0, 1]$ rather than on the power set of a probability space. The computational convenience of ROC curves is also shown in the context of the divergences (distance functions between probability measures) that can be associated to families of distortion risk measures. An exponential reduction in computation time is achieved with respect to a naive approach towards the calculation of a divergence of this type between two given probability measures.

The construction of distortion functions from ROC curves associated to hypothesis testing problems does not lead to a new class of distortion risk measures; as discussed in Remark 3.12 below, any concave distortion function can be obtained as an ROC curve. However, the connection to hypothesis testing problems via ROC curves provides a statistical interpretation of distortion risk measures. Some distortion functions are related to quite standard and well known testing problems, while others relate to more involved problems, as shown by a series of examples in Section 3. This helps create structure in the wide landscape of distortion functions.

Reesor and McLeish (2003) construct distortions via relative entropy minimization subject to moment constraints. This approach is quite different from the one taken here. More closely related to the present work is a paper by Tsukahara (2009), in which one-parameter families of distortion functions are formed by means of a translation operation. This construction can be viewed as a special case of the construction used here; see Example 3.10 below. Tsukahara (2009) does not discuss the interpretation via hypothesis testing that is central to the present paper.

The paper is organized as follows. In Section 2, a brief introduction is given to ROC curves, with inclusion of basic facts that are well known but that are stated here for easy reference. Relations between ROC curves and distortion functions are discussed in Section 3. The concept of distortion divergence is introduced in Section 4, and conclusions follow in Section 5.

2 ROC curves

Often, in the context of hypothesis testing, one-parameter families of tests are considered in which any particular desired significance level can be reached by choosing an appropriate value of the parameter. In the case of testing a simple null against a simple alternative, the associated *receiver operating characteristic*[1] for such a family of tests is defined as the plot of the probability of rejection of the null under the alternative hypothesis against the probability of rejection of the null under the null hypothesis. The term *ROC curve* will be used in this paper to refer to the *optimal* receiver operating characteristic corresponding to a pair of probability measures (P_0, P_1). The ROC curve in this sense may be introduced more formally as follows.

Throughout the paper, we work within a setting in which a probability space $(\Omega, \mathcal{F}, P_0)$ is given, together with a second probability measure P_1 that is absolutely continuous with respect to P_0. A *test* is a random variable taking values in $[0, 1]$, interpreted as the probability of rejecting the null hypothesis. The set of tests is denoted by $L(\Omega, \mathcal{F}; [0, 1])$. Define the *ROC set* within the unit square $[0, 1] \times [0, 1]$ by

$$\mathrm{ROC}(P_0, P_1) = \{(E_0\varphi, E_1\varphi) \mid \varphi \in L(\Omega, \mathcal{F}; [0, 1])\}$$

where E_0 and E_1 denote expectation under P_0 and under P_1 respectively. The quantity $E_0\varphi$ is the size of the test φ, and $E_1\varphi$ is its power. For discussions of this set in the literature, see for instance Dudewicz and Mishra (1988, Section 9.4), where it appears under the name "set of risk points", Lehmann and Romano (2005, Section 3.2), or, for an optimization perspective on the discrete case, Boyd and Vandenberghe (2004, Subsection 7.3.5). The ROC set is convex, because it is the image of a convex set under a linear mapping. It contains the diagonal, since one can take $\varphi = \alpha$ where $\alpha \in [0, 1]$. Now define the *ROC curve* as the function that describes the upper boundary of the ROC set, i.e.,

$$\mathrm{ROC}_{P_0,P_1}(u) = \sup\{E_1\varphi \mid \varphi \in L(\Omega, \mathcal{F}; [0, 1]),\ E_0\varphi = u\}, \qquad 0 \le u \le 1. \tag{2.1}$$

[1]This commonly used phrase appears to be a blend of the term "operating characteristic" that was used by Wald (1945, p. 162) and the term "receiver operator characteristic" that was used in Britain in the Second World War to describe the performance of operators of radar receivers under various settings of the signal gain.

By the Neyman-Pearson lemma (cf. Lemma 2.4 below), the supremum is achieved for every $u \in [0,1]$. The ROC curve describes the maximal power that can be achieved by tests of P_0 against P_1 at any given significance level; as such, it gives an indication of the degree to which the probability measures P_0 and P_1 are different.

The following proposition states a few basic properties of the ROC curve. For proofs, see for instance Lehmann and Romano (2005, Section 3.2).

Proposition 2.1. *The function* ROC_{P_0,P_1} *is nondecreasing, concave, maps 0 to 0 and 1 to 1, and satisfies* $\mathrm{ROC}_{P_0,P_1}(u) \geq u$ *for all* $u \in [0,1]$.

Concerning the behavior at 0, the following statement can be made.

Proposition 2.2. *The quotient* $\mathrm{ROC}_{P_0,P_1}(u)/u$ *has a (finite or infinite) limit as* u *tends to* 0, *and we have*

$$\lim_{u\downarrow 0} \frac{\mathrm{ROC}_{P_0,P_1}(u)}{u} = \sup_{0<u\leq 1} \frac{\mathrm{ROC}_{P_0,P_1}(u)}{u} = \operatorname{ess\,sup}_{P_0} \frac{dP_1}{dP_0}\,. \tag{2.2}$$

Proof. The concavity of the ROC curve implies that $\mathrm{ROC}_{P_0,P_1}(u_1)/u_1 \geq \mathrm{ROC}_{P_0,P_1}(u_2)/u_2$ for $0 < u_1 \leq u_2 \leq 1$. This assures the existence of the limit in (2.2) and proves the equality of the limit and the supremum. Suppose now that $b \in \mathbb{R}$ is a P_0-essential upper bound of the Radon-Nikodym derivative $h := dP_1/dP_0$, i.e., $P_0(h \leq b) = 1$. For any test φ, we have $E_1\varphi = \int h\varphi\, dP_0 \leq b \int \varphi\, dP_0 = bE_0\varphi$, so that b is also an upper bound for the quotient $\mathrm{ROC}_{P_0,P_1}(u)/u$. This shows that $\sup_{0<u\leq 1} \mathrm{ROC}_{P_0,P_1}(u)/u \leq \operatorname{ess\,sup}_{P_0} h$. To prove that strict inequality cannot hold, assume the contrary and take a such that $\sup_{0<u\leq 1} \mathrm{ROC}_{P_0,P_1}(u)/u < a < \operatorname{ess\,sup}_{P_0} h$. In particular, we then have $P_0(h > a) > 0$. Take $\varphi = \mathbb{1}_{h>a}$. Then $E_0\varphi > 0$, and, since $h\varphi \geq a\varphi$, $E_1\varphi = \int h\varphi\, dP_0 \geq a \int \varphi\, dP_0 = aE_0\varphi$. Consequently, for $u = E_0\varphi > 0$, we have $\mathrm{ROC}_{P_0,P_1}(u)/u \geq a$, and we arrive at a contradiction. □

Remark 2.3. The existence of the limit in (2.2) means that the ROC curve has a one-sided derivative at $u = 0$. Therefore, one might also write

$$\mathrm{ROC}'_{P_0,P_1}(0) = \operatorname{ess\,sup}_{P_0} \frac{dP_1}{dP_0}\,. \tag{2.3}$$

In a similar way, one can prove

$$\mathrm{ROC}'_{P_0,P_1}(1) = \operatorname{ess\,inf}_{P_0} \frac{dP_1}{dP_0}\,. \tag{2.4}$$

For easy reference, the Neyman-Pearson lemma is stated here in the form in which it will be needed below, with a brief reminder of the proof.

Lemma 2.4. *Let a probability space* $(\Omega, \mathcal{F}, P_0)$ *be given, and let* P_1 *be a probability measure on* $(\Omega, \mathcal{F})$ *with* $P_1 \ll P_0$. *Let* $h : \Omega \to [0, \infty)$ *denote a version of the Radon-Nikodym derivative of* P_1 *with respect to* P_0, *and let* E_0 *and* E_1 *denote the expectation operators corresponding to* P_0 *and* P_1 *respectively. Furthermore, let a number* $k \geq 0$ *and a number* $\theta \in [0,1]$ *be given. Define a*

$[0,1]$-valued random variable φ as follows:

$$\varphi(\omega) = \begin{cases} 1 & \text{if } h(\omega) > k \\ \theta & \text{if } h(\omega) = k \\ 0 & \text{if } h(\omega) < k. \end{cases} \tag{2.5}$$

Then, for $\psi \in L(\Omega, \mathcal{F}; [0,1])$, the following implication holds:

$$E_0\psi \le E_0\varphi \;\Rightarrow\; E_1\psi \le E_1\varphi. \tag{2.6}$$

Proof. Note that it is sufficient to show that $E_1[\psi - \varphi] \le kE_0[\psi - \varphi]$, and that this inequality may also be written as $E_0[(h-k)(\psi-\varphi)] \le 0$. But the inequality $(h-k)(\psi-\varphi) \le 0$ already holds on a pointwise basis, as is verified by separate consideration of the three cases in (2.5). □

Remark 2.5. For every $0 < u \le 1$, one can find a number $k \ge 0$ such that $P_0(h \ge k) \ge u \ge P_0(h > k)$; this implies that there exists a test φ of the form (2.5) such that $E_0\varphi = u$. In other words, the Neyman-Pearson tests (together with the point $(0,0)$, which corresponds to the test $\varphi = 0$) are sufficient to trace out the ROC curve.

The Neyman-Pearson lemma can be used to establish an additional basic fact concerning the ROC curve.

Proposition 2.6. *The function* ROC_{P_0,P_1} *is continuous.*

Proof. Continuity of ROC_{P_0,P_1} on the open interval $(0,1)$ follows from a general result on convex functions; see for instance Rockafellar (1997, Thm. 10.1). Continuity at the point 1 follows from the inequalities $u \le \mathrm{ROC}_{P_0,P_1}(u) \le \mathrm{ROC}_{P_0,P_1}(1) = 1$ which hold for $0 \le u \le 1$. It remains to establish the continuity at $u = 0$. The Neyman-Pearson lemma shows that all points on the ROC curve can be written in the form $(E_0\varphi, E_1\varphi)$ where $\varphi = \mathbb{1}_{h>k} + \theta\mathbb{1}_{h=k}$ with $0 \le \theta \le 1$ and $k \ge 0$. Because we can write

$$\big(E_0\varphi, E_1\varphi\big) = (1-\theta)\big(P_0(h>k), P_1(h>k)\big) + \theta\big(P_0(h \ge k), P_1(h \ge k)\big)$$

this shows that the ROC curve lies in the convex hull of the points $\{\big(P_0(A), P_1(A)\big) \mid A \in \mathcal{F}\}$. Consequently, to show the continuity of the ROC curve at $u = 0$ it is sufficient to prove that for every $\varepsilon > 0$, there exists $\delta > 0$ such that $P_0(A) < \delta$ implies $P_1(A) < \varepsilon$. Using the Borel-Cantelli lemma, one can show that this follows from the absolute continuity of P_1 with respect to P_0; see for instance Billingsley (1986, p. 443). □

Remark 2.7. It can be shown (compare Cor. 3.4 below) that the ROC set is in fact equal to the convex hull of the set $\{\big(P_0(A), P_1(A)\big) \mid A \in \mathcal{F}\}$. In the finite case, this set might be described as the scatter plot of the two functions from $\mathcal{F}$ to $[0,1]$ that are given by the measures P_0 and P_1.

If the two probability measures P_0 and P_1 are equivalent, then one can also define the ROC curve of P_1 versus P_0. The function $\hat{g}(u) := \mathrm{ROC}_{P_1,P_0}(u)$ is in general not the same as the function $g(u) := \mathrm{ROC}_{P_0,P_1}(u)$. The relation between the two functions can be described as follows.

Proposition 2.8. *If the measure P_1 is equivalent to P_0, then the ROC curve of P_0 versus P_1, $g(u) := \mathrm{ROC}_{P_0,P_1}(u)$, is strictly increasing. Let its inverse be denoted by $g^{-1} : [0,1] \to [0,1]$. The ROC curve of P_1 versus P_0, $\hat{g}(u) := \mathrm{ROC}_{P_1,P_0}(u)$, is related to the ROC curve of P_0 versus P_1 by*

$$\hat{g}(u) = 1 - g^{-1}(1-u). \tag{2.7}$$

Proof. The ROC curve is nondecreasing; therefore, in order to prove that the curve is strictly increasing, it is enough to show that the relation $g(u_1) = g(u_2)$ can only hold when $u_1 = u_2$. By the Neyman-Pearson lemma, this follows if it can be shown that the implication

$$E_1\varphi = E_1\varphi' \quad \Rightarrow \quad E_0\varphi = E_0\varphi' \tag{2.8}$$

holds for all random variables φ and φ' of the form (2.5). Therefore, let φ and φ' be defined by

$$\varphi = \mathbb{1}_{h>k} + \theta\mathbb{1}_{h=k}\,, \qquad \varphi' = \mathbb{1}_{h>k'} + \theta'\mathbb{1}_{h=k'}$$

with $0 \le k \le k'$ (without loss of generality), $0 \le \theta \le 1$, and $0 \le \theta' \le 1$. Note that

$$\begin{aligned} E_1\varphi - E_1\varphi' &= P_1(h > k) + \theta P_1(h = k) - \big(P_1(h > k') + \theta' P_1(h = k')\big) \\ &= P_1(k < h < k') + P_1(h = k') + P_1(h > k') + \theta P_1(h = k) \\ &\qquad\qquad\qquad\qquad - \big(P_1(h > k') + \theta' P_1(h = k')\big) \\ &= P_1(k < h < k') + (1-\theta')P_1(h = k') + \theta P_1(h = k). \end{aligned} \tag{2.9}$$

Since the three terms at the right hand side are all nonnegative, the equality $E_1\varphi = E_1\varphi'$ implies that each of the three terms is zero. By the assumed equivalence of P_1 and P_0, it follows from $P_1(k < h < k') = 0$ that $P_0(k < h < k') = 0$. The relation $(1-\theta')P_1(h = k') = 0$ implies that $\theta' = 1$ or $P_1(h = k') = 0$; in both cases, it follows that $(1-\theta')P_0(h = k') = 0$. Likewise, $\theta P_1(h = k) = 0$ implies $\theta P_0(h = k) = 0$. Using the relation analogous to (2.9) for P_0 instead of P_1, it follows that $E_0\varphi = E_0\varphi'$. We can conclude that the ROC curve of P_0 versus P_1 is strictly increasing, so that the inverse function is well-defined. The expression (2.7) then follows directly from the definitions. □

Under the mapping defined by (2.7), a point (u, v) lies on the reversed curve $\hat{g}$ if and only if $(1-v, 1-u)$ lies on the original curve g. The mapping from (u, v) to $(1-v, 1-u)$ can be described most directly as reflection around the cross-diagonal in the unit square (i.e. the diagonal that runs from $(0, 1)$ to $(1, 0)$). Indeed, points that satisfy $u + v = 1$ are invariant under the mapping, the difference $u - v$ is preserved, and applying the mapping twice produces the identity. For a graphical example, see Fig. 1 in Section 3.

Example 2.9. The simplest example of an ROC curve arises from the null hypothesis $\theta = p$ against the alternative $\theta = q$ for a Bernoulli random variable with success probability θ. If $q \ge p$, the ROC curve is

$$\mathrm{ROC}_{P,Q}(u) = \frac{q}{p}\min(u, p) + \frac{1-q}{1-p}\max(u-p, 0)$$

and in the opposite case the curve is given by

$$\mathrm{ROC}_{P,Q}(u) = \frac{1-q}{1-p}\min(u, 1-p) + \frac{q}{p}\max(u-1+p, 0).$$

3 Distortion functions

A *distortion function* is a nondecreasing, continuous and concave function $g : [0,1] \to [0,1]$ which satisfies $g(0) = 0$ and $g(1) = 1$. Given such a function, the corresponding *distortion risk measure* ρ_g is defined by (Wang, 1996)

$$\rho_g(X) = -\int_{-\infty}^{0} \bigl(1 - g(\bar{F}(x))\bigr)\,dx + \int_0^\infty g(\bar{F}(x))\,dx$$

where $\bar{F}$ is the decumulative distribution function of the random variable X (i.e., $\bar{F}(x) = P(X > x)$). The above expression represents in fact the Choquet integral (Choquet, 1954) of X with respect to the capacity defined by the non-additive set function $c_g(A) = g(P(A))$. An alternative expression for the distortion risk measure is therefore given by (Denneberg, 1994)

$$\rho_g(X) = \inf_{Q\in\mathcal{Q}} E^Q X \tag{3.1}$$

where

$$\mathcal{Q} = \{Q \mid Q(A) \le g(P(A)) \text{ for all } A \in \mathcal{F}\}. \tag{3.2}$$

The following proposition on the relationship between ROC curves and distortion functions is key to the development of this paper.

Proposition 3.1. *Consider a probability space $(\Omega, \mathcal{F}, P_0)$. Let $g : [0,1] \to [0,1]$ be concave. A probability measure $P_1 \ll P_0$ satisfies*

$$P_1(A) \le g(P_0(A)) \qquad \text{for all } A \in \mathcal{F} \tag{3.3}$$

if and only if

$$\mathrm{ROC}_{P_0,P_1}(u) \le g(u) \qquad \text{for all } u \in [0,1]. \tag{3.4}$$

Proof. By definition of the ROC curve, the relation (3.4) is equivalent to the condition

$$E_1\varphi \le g(E_0\varphi) \quad \text{for all } \varphi \in L(\Omega, \mathcal{F}; [0,1]). \tag{3.5}$$

First suppose that (3.4) holds. Then (3.3) follows by applying (3.5) to random variables of the form $\varphi = \mathbb{1}_A$ with $A \in \mathcal{F}$. For the converse, suppose that (3.3) holds. By the Neyman-Pearson lemma and Remark 2.5 above, the relation (3.5) (and hence (3.4)) follows if the inequality $E_1\varphi \le g(E_0\varphi)$ holds for all random variables of the form (2.5). For given $k \ge 0$ and $0 \le \theta \le 1$, define $\varphi = \mathbb{1}_{h>k} + \theta\mathbb{1}_{h=k}$. We have

$$E_i\varphi = P_i(h > k) + \theta P_i(h = k) = \theta P_i(h \ge k) + (1-\theta)P_i(h > k) =: \theta a_i + (1-\theta)b_i$$

for $i = 0, 1$. The condition (3.3) implies $a_1 \leq g(a_0)$ and $b_1 \leq g(b_0)$. By concavity of g, it follows that $E_1\varphi = \theta a_1 + (1-\theta)b_1 \leq \theta g(a_0) + (1-\theta)g(b_0) \leq g\big(\theta a_0 + (1-\theta)b_0\big) = g(E_0\varphi)$. □

Remark 3.2. The condition (3.4) may be compared to the condition for "statistical indistinguishability" proposed by Balter and Pelsser (2017). In the notation of the present paper, the condition of Balter and Pelsser (2017) can be written as $\mathrm{ROC}(\alpha) \leq 1 - \beta$, where the numbers α and β are fixed. According to the authors, these numbers can be selected, for instance, on the basis of usual conventions in statistics ($\alpha = 0.05$, $\beta = 0.2$). Such a choice of particular levels of significance and power is avoided in (3.4). On the other hand, the criterion (3.4) calls for the selection of a distortion function. Note, however, that the relation (3.4) can be used in reverse (taking the left hand side as given, rather than the right hand side) to replace the selection of a distortion function by the selection of a particular testing problem, which is then effectively given the role of a benchmark. This strategy is applied below.

As a corollary to the proposition above, another description of the ROC curve can be given as follows. First, introduce a definition.

Definition 3.3. Given a probability space $(\Omega, \mathcal{F}, P_0)$ and a probability measure $P_1 \ll P_0$, the function $\mathrm{roc}_{P_0,P_1} : [0,1] \to [0,1]$ is defined by

$$\mathrm{roc}_{P_0,P_1}(u) = \sup\{P_1(A) \mid P_0(A) \leq u\}. \tag{3.6}$$

The function roc_{P_0,P_1} is not always concave; for instance, when Ω is finite, it is in fact a piecewise constant function.

Corollary 3.4. *The function* ROC_{P_0,P_1} *is the least concave majorant of* roc_{P_0,P_1}.

Proof. It was already noted in Prop. 2.2 that the function ROC_{P_0,P_1} is concave. To show that it is a majorant of roc_{P_0,P_1}, take $u \in [0,1]$ and let $A \in \mathcal{F}$ be such that $P_0(A) \leq u$. Define $\varphi = \mathbb{1}_A$. Using the fact that the ROC curve is nondecreasing, we can write

$$\mathrm{ROC}_{P_0,P_1}(u) \geq \mathrm{ROC}_{P_0,P_1}(P_0(A)) = \mathrm{ROC}_{P_0,P_1}(E_0\varphi) \geq E_1\varphi = P_1(A)$$

where the final inequality holds by definition of the ROC curve. This shows that $\mathrm{ROC}_{P_0,P_1}(u)$ is an upper bound for the set $\{P_1(A) \mid P_0(A) \leq u\}$, and hence that $\mathrm{ROC}_{P_0,P_1}(u) \geq \mathrm{roc}_{P_0,P_1}(u)$.

To show that ROC_{P_0,P_1} is the least concave majorant of roc_{P_0,P_1}, let g be an arbitrary concave function satisfying $g(u) \geq \mathrm{roc}_{P_0,P_1}(u)$ for all u. Take $A \in \mathcal{F}$ and write $u = P_0(A)$. By definition of the function roc_{P_0,P_1}, we have

$$P_1(A) \leq \mathrm{roc}_{P_0,P_1}(u) \leq g(u) = g(P_0(A)).$$

Since this holds for all $A \in \mathcal{F}$, it follows by Prop. 3.1 that $\mathrm{ROC}_{P_0,P_1} \leq g$. □

As already argued above, distortion functions can be understood as a way of defining a collection of probability measures that are close to a given reference measure in a specific sense. The ROC curve likewise expresses a notion of closeness between two probability measures. Prop. 3.1

above indicates the connection between the two viewpoints. The proposition suggests that a distortion curve that is calibrated on a particular pair of probability measures (P_0, P_1), so to say, can be constructed by taking the ROC curve of the pair as the distortion:

$$g(u) = \mathrm{ROC}_{P_0,P_1}(u) \qquad (0 \leq u \leq 1). \tag{3.7}$$

It should be emphasized that there is not necessarily any relationship between, on the one hand, the probability space on which P_0 and P_1 are defined, and, on the other hand, the probability space on which the reference measure P is constructed; nor does there need to be any relation between P and P_0. The class of probability measures defined by (3.2) through the combination of the reference measure P and the distortion function defined in (3.7) might be described as the class of probability measures that are at least as difficult to distinguish from P as P_1 is from P_0. In this sense, the pair (P_0, P_1) is used as a benchmark. As a specific example, suppose that risk is measured by a distortion function from Wang's family, with parameter value $\eta = 1$ (parametrization as in Table 1). Then the statement that a given risk X is acceptable, in the sense of Artzner et al. (1999), could be phrased as (see Example 3.5 below) "the expected return on the position does not turn negative, even when it is allowed to distort the probabilities of events so much that, statistically speaking, the distorted measure becomes as easy to distinguish from the reference measure as the $N(1,1)$ distribution is from the $N(0,1)$ distribution". While no one-dimensional criterion can do full justice to the multi-faceted nature of risk, such a statement may be helpful for researchers and practitioners to get a feeling for the conservativeness of parameter choices.

Many of the distortion risk measures that have been proposed in the literature are actually defined as members of one-parameter *families* of risk measures, corresponding to one-parameter families of distortion functions. In this way, it is possible for the user to express various levels of conservatism. Examples are provided in Table 1, where the parametrization, in terms of a parameter $\eta \geq 0$, is always chosen in such a way that $\eta = 0$ corresponds to absence of distortion, and, for fixed $u \in [0,1]$, the function $\eta \mapsto g_\eta(u)$ is nondecreasing. This expresses that larger values of η are associated with larger distortions and hence with more conservatism, so that η can be viewed as a parameter of risk aversion.[2] It should be noted that the choice of a parametrization is to some extent arbitrary; the same family of distortion functions may be described by different (monotonically related) parametrizations.

Like distortion functions, probability distributions are usually introduced as parametrized families. The relation (3.7) can be used to create families of distortion functions from families of probability distributions. Some of these family connections are shown in the following examples.

Example 3.5. Consider the testing problem $H_0 : X \sim N(0,1)$ vs. $H_1 : X \sim N(\mu,1)$, where μ is a given positive number. To construct the ROC curve, it suffices to consider tests of the form $\varphi = \mathbb{1}_{X \geq c}$ with $c \in \mathbb{R}$. For such tests, we have

$$E_0\varphi = \Phi(-c), \qquad E_1\varphi = \Phi(\mu - c)$$

[2]Terminology here follows the usage of "risk measure" in Artzner et al. (1999). The set defined in (3.2) can be thought of as representing model uncertainty, and from that point of view the parameter η may be described as indicating ambiguity aversion (Ellsberg, 1961; Gilboa and Schmeidler, 1989; Hansen and Sargent, 2008).

name tag	$g_\eta(u)$	reference
CVaR	$\min\big((1+\eta)u, 1\big)$	Artzner et al. (1999)
Wang	$\Phi\big(\Phi^{-1}(u)+\eta)\big)$	Wang (2000)
PH	$u^{1/(1+\eta)}$	Wang (1995)
dual power	$1-(1-u)^{1+\eta}$	Wirch and Hardy (1999)
minmaxvar	$1-(1-u^{1/(1+\eta)})^{1+\eta}$	Cherny and Madan (2009)

Table 1 – One-parameter families of distortion functions. Parametrization has been adapted so that the range of the parameter η is $[0, \infty)$, and $g_0(u) = u$ for all u.

where Φ is the standard normal cumulative distribution function. The corresponding ROC curve is given by

$$g_N(u;\mu) = \Phi\big(\Phi^{-1}(u)+\mu\big) \qquad (\mu \geq 0).$$

Taking μ as a parameter, one is led to the family of distortion functions proposed by Wang (2000).

Example 3.6. Consider the testing problem $H_0\colon X \sim Exp(1)$ vs. $H_1 : X \sim Exp(\lambda)$, where $\lambda \in (0, 1]$ is given, and $Exp(\lambda)$ refers to the exponential distribution with parameter λ. The ROC curve is traced out by tests of the form $\varphi = \mathbb{1}_{X\geq c}$ with $c \geq 0$. For such tests, we have

$$E_0\varphi = e^{-c}, \qquad E_1\varphi = e^{-\lambda c}$$

so that the corresponding ROC curve is

$$g_{Exp}(u;\lambda) = u^\lambda \qquad (0 < \lambda \leq 1).$$

This is, up to reparametrization, the PH (proportional hazards) family of distortion functions (Wang, 1995).

Example 3.7. Consider the testing problem $H_0\colon X \sim U(0, 1)$ vs. $H_1 : X \sim U(0, a)$, where $a \in (0, 1]$ is given, and $U(0, a)$ refers to the uniform distribution on the interval $[0, a]$. In this case, P_1 is absolutely continuous with respect to P_0, but not equivalent to P_0. Under the canonical choices $\Omega = [0, 1]$ and $X(\omega) = \omega$, and with $h = a^{-1}\mathbb{1}_{[0,a]}$ as the Radon-Nikodym derivative, we have $\{h = 0\} = (a, 1]$ and $\{h = 1/a\} = [0, a]$. The Neyman-Pearson tests are of one of the two forms $\theta\mathbb{1}_{[0,a]}$ and $\mathbb{1}_{[0,a]} + \theta\mathbb{1}_{(a,1]}$, with $0 \leq \theta \leq 1$. For $\varphi = \theta\mathbb{1}_{[0,a]}$, we have $E_0\varphi = \theta a$ and $E_1\varphi = \theta$. For $\varphi = \mathbb{1}_{[0,a]} + \theta\mathbb{1}_{(a,1]}$, we have $E_0\varphi = a + \theta(1 - a)$ and $E_1\varphi = 1$. The corresponding ROC curve is

$$g_U(u;a) = \min(u/a, 1) \qquad (a \leq 1).$$

This is the CVaR family.

Example 3.8. Consider the testing problem $H_0\colon X \sim Ber(p)$ vs. $H_1 : X = 1$, where $Ber(p)$ refers to the Bernoulli distribution that assigns probability p to the outcome 1 and probability $1 - p$ to the outcome 0, and where the alternative is a degenerate random variable. The function roc_{P_0,P_1} is given by $\mathrm{roc}_{P_0,P_1}(u) = 0$ for $0 \leq u < p$ and $\mathrm{roc}_{P_0,P_1}(u) = 1$ for $p \leq u \leq 1$. Its smallest concave majorant is the function $g_B(u;p) = \min(u/p, 1)$. Again we arrive at the CVaR family.

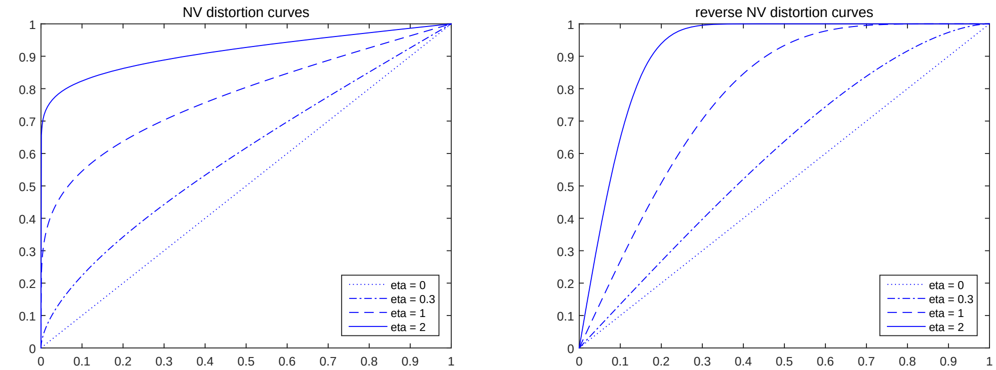

Figure 1 – Distortion curves from the normal-variance family and its reverse.

Example 3.9. Consider the testing problem $H_0 : X \sim N(0,1)$ vs. $H_1 : X \sim N(0,\sigma^2)$ with $\sigma \geq 1$. In this case, tests that determine the ROC curve are of the form $\varphi = \mathbb{1}_{|X|\geq c}$ with $c \geq 0$. For these tests, we have

$$E_0\varphi = 2\Phi(-c), \qquad E_1\varphi = 2\Phi(-c/\sigma).$$

The ROC curve is

$$g_{NV}(u;\sigma) = 2\Phi\big(\Phi^{-1}(\tfrac{1}{2}u)/\sigma\big) \qquad (\sigma \geq 1)$$

where the subscript refers to the normal-variance testing problem from which the family is derived. A reparametrization on $[0,\infty)$ as in Table 1 can be given as follows (abusing notation by not changing the function symbol):

$$g_{NV}(u;\eta) = 2\Phi\big(e^{-\eta}\,\Phi^{-1}(\tfrac{1}{2}u)\big) \qquad (\eta \geq 0). \tag{3.8}$$

As far as the author is aware, this family of distortion functions has not appeared in the literature before. By reversing the role of the null and the alternative in the above testing problem, or equivalently by considering the same testing problem but now under the assumption $\sigma^2 \leq 1$ rather than $\sigma^2 \geq 1$, one obtains the reverse family

$$\hat{g}_{NV}(u;\eta) = 1 - 2\Phi\big(e^{\eta}\,\Phi^{-1}(\tfrac{1}{2}(1-u))\big) = 2\Phi\big(e^{\eta}\,\Phi^{-1}(\tfrac{1}{2}+\tfrac{1}{2}u)\big) - 1 \qquad (\eta \geq 0). \tag{3.9}$$

A few members of the NV family and its reverse are shown in Fig. 1. The reverse NV distortions show approximately linear behavior for small values of the argument u, and take values very close to 1 for larger arguments when the risk aversion parameter η is not too small. These distortions can therefore potentially be used as smoothed versions of CVaR.

Example 3.10. More generally than in Example 3.5 above, one can consider a testing problem of the form $H_0 : X \sim F$ vs. $H_1 : X \sim F_m$ where F is a given strictly increasing cumulative distribution function, and F_m, with $m \in \mathbb{R}$, represents a shifted version of F (i.e., $F_m(x) = F(x+m)$). Under the condition that the distribution function F is *strongly unimodal* (meaning that it has a density f, and $\log f$ is concave), the function $h(x) := f(x+m)/f(x)$ is monotonic; see for instance Arnold et al. (1987). The ROC curve for $m \geq 0$ is therefore determined by tests of the form $\mathbb{1}_{X\leq c}$. The corresponding ROC curve is

$$g_F(u;m) = F\big(F^{-1}(u) + m\big) \qquad (m \geq 0).$$

One-parameter families of distortion functions that are generated in this way were studied by Tsukahara (2009), without the motivation from hypothesis testing that is used here. As a particular case, it is shown in the cited paper that the PH family of distortion functions can be obtained from a translation family of Gumbel distributions. As shown above in Example 3.6, the PH family can also be constructed from the exponential distribution, using a family that is not of translation type.

Remark 3.11. In all cases where the null and the alternative define equivalent probability measures, one may also consider the testing problem in which the roles of the null and the alternative are reversed. For instance, the reverse obtained in this way from the PH family is the dual power family (cf. Prop. 2.8). Wang's distortion function is invariant under reversion; this reflects the symmetry of the testing problem in Example 3.5.

Remark 3.12. Given a distortion function g, one can take this function as the cumulative distribution function of a random variable with values in the unit interval. One can then consider the problem of testing $H_0\colon X \sim U(0,1)$ versus $H_1\colon X \sim g$. The concavity of g, which in this paper is part of the definition of distortion functions, implies that the measure defined by g has a nonincreasing Radon-Nikodym derivative with respect to the uniform measure. Consequently, the ROC curve is determined by Neyman-Pearson tests of the form $\varphi = \mathbb{1}_{X \leq c}$ with $0 \leq c \leq 1$. For such a test we have $E_0\varphi = c$ and $E_1\varphi = g(c)$; in other words, the ROC curve is given by the function g. This construction provides a "canonical" way to find a testing problem whose ROC curve is a given distortion function. This testing problem may however not be the most natural one associated with the given distortion.

To apply the criterion (3.4), one needs to compute the ROC curve for a given pair of probability measures. In some cases, as demonstrated above, the ROC curve can be calculated analytically, but in other cases a numerical computation will be needed. Consider two probability measures which will now be called P and Q to prevent overloading of numerical subscripts, and assume that these measures have been obtain by discretization or were already finitely discrete to begin with. Write $\Omega = \{\omega_1, \ldots, \omega_n\}$. Absolute continuity of Q with respect to P is reflected in the assumptions $p_i := P(\omega_i) > 0$ and $q_i := Q(\omega_i) \geq 0$ for all i. As a first step, renumber the sample points, if necessary, such that

$$\frac{q_1}{p_1} \geq \frac{q_2}{p_2} \geq \cdots \geq \frac{q_n}{p_n}\,. \tag{3.10}$$

Define the sets A_i $(i = 0, \ldots, n)$ by

$$A_0 = \emptyset, \qquad A_i = \{\omega_1, \ldots, \omega_i\} \qquad (i = 1, \ldots, n). \tag{3.11}$$

We have $P(A_i) = \sum_{j=1}^{i} p_j$, $Q(A_i) = \sum_{j=1}^{i} q_j$ for $i = 0, \ldots, n$. From the ordering (3.10) it follows that we also have the ordering

$$\frac{Q(A_1)}{P(A_1)} \geq \frac{Q(A_2)}{P(A_2)} \geq \cdots \geq \frac{Q(A_n)}{P(A_n)} \tag{3.12}$$

since, for each $i = 1, \dots, n-1$,

$$Q(A_i) = \sum_{j=1}^{i} q_j = \sum_{j=1}^{i} \frac{q_j}{p_j}\, p_j \geq \frac{q_{i+1}}{p_{i+1}} \sum_{j=1}^{i} p_j = \frac{Q(A_{i+1}) - Q(A_i)}{P(A_{i+1}) - P(A_i)}\, P(A_i).$$

To reduce the computation of the ROC curve to the computation of a finite set of numbers, one can make use of the following proposition.

Proposition 3.13. *The ROC curve for a pair of probability measures P and Q defined on a finite sample space Ω, with $P(\omega) > 0$ for all $\omega \in \Omega$, is the piecewise linear function defined by the interpolation points $\big(P(A_i), Q(A_i)\big)$ $(i = 0, \dots, n)$, where the sets A_i are defined by (3.10) and (3.11).*

Proof. The ordering (3.10) can be written in more detail as

$$\frac{q_1}{p_1} = \cdots = \frac{q_{i_1}}{p_{i_1}} > \frac{q_{i_1+1}}{p_{i_1+1}} = \cdots = \frac{q_{i_1+i_2}}{p_{i_1+i_2}} > \cdots > \frac{q_{i_1+\cdots+i_{k-1}+1}}{p_{i_1+\cdots+i_{k-1}+1}} = \cdots = \frac{q_{i_1+\cdots+i_k}}{p_{i_1+\cdots+i_k}} \tag{3.13}$$

where $i_1 + \cdots + i_k = n$. Define $B_0 = \emptyset$, $B_j = A_{i_1+\cdots+i_j}$, $j = 1, \dots, k$. For brevity, write $u_j = P(B_j)$ and $v_j = Q(B_j)$ $(j = 0, \dots, k)$. Take $u \in (0,1)$ and let j be such that $u_j < u \leq u_{j+1}$. Choose $\lambda \in (0,1]$ such that $u = (1-\lambda)u_j + \lambda u_{j+1}$. Define a test φ by

$$\varphi = (1-\lambda)\mathbb{1}_{B_j} + \lambda \mathbb{1}_{B_{j+1}}.$$

Note that this is a Neyman-Pearson test with parameters $k = q_{j+1}/p_{j+1}$ and $\theta = \lambda$. We have

$$E^P\varphi = (1-\lambda)P(B_j) + \lambda P(B_{j+1}) = (1-\lambda)u_j + \lambda u_{j+1} = u.$$

For $A \in \mathcal{F}$ with $P(A) \leq u$, we have $E^P\mathbb{1}_A \leq E^P\varphi$ and hence, by the Neyman-Pearson lemma,

$$Q(A) = E^Q\mathbb{1}_A \leq E^Q\varphi = (1-\lambda)Q(B_j) + \lambda Q(B_{j+1}) = (1-\lambda)v_j + \lambda v_{j+1}.$$

This implies that the curve $\mathrm{roc}_{P,Q}$ is majorized by the piecewise linear interpolant of the points (u_j, v_j). Since, on the other hand, the inequality $\mathrm{roc}_{P,Q}(u_i) \geq v_i$ for $i = 0, \dots, n$ follows from application of the definition of the roc function to the sets B_j, and since concavity of the piecewise linear interpolant is a consequence of the inequalities (3.12), we can conclude that the piecewise linear interpolant is the smallest concave majorant of $\mathrm{roc}_{P,Q}$, and hence (by Cor. 3.4) is equal to $\mathrm{ROC}_{P,Q}$.

It remains to show that the piecewise linear interpolant of the points $\big(P(B_j), Q(B_j)\big)$ $(j = 0, \dots, k)$ is the same as the piecewise linear interpolant of the points $\big(P(A_i), Q(A_i)\big)$ $(i = 1, \dots, n)$. This follows from the equalities in (3.13). □

To determine the ROC curve, it therefore suffices to compute the points $(P(A_i), Q(A_i))$ for $i = 1, \dots, n$. Due to the concavity of the distortion function g, verification of the inequality in (3.4) for all $u \in [0,1]$ requires only checking that the inequality holds at the interpolation points $\big(P(A_i), Q(A_i)\big)$. This can be readily implemented in standard software.

4 Distortion divergence

Divergences are distance measures (not necessarily metrics) between probability distributions. They have been used in many applications including optimization, data processing, uncertainty quantification, and robustness analysis; see for instance Bregman (1967); Goll and Rüschendorf (2001); Basseville (2013); Postek et al. (2016) to mention just a few. Broad classes of divergences include the so called f-divergences and the Bregman divergences, which are both generated from the choice of a convex function. Likewise, a broad class of divergences can be generated from the choice of a family of distortion functions. The construction proceeds as follows.

Consider a one-parameter family of distortion functions $(g_\eta)_{\eta\geq 0}$, parametrized in such a way that the following conditions are satisfied.

Assumption 4.1. The family of distortion functions g_η $(\eta \geq 0)$ satisfies the following properties:

(i) $g_0(u) = u$ for all $u \in [0,1]$ (i.e., $\eta = 0$ corresponds to no distortion)

(ii) for each fixed $u \in [0,1]$, the function $\eta \mapsto g_\eta(u)$ is continuous and nondecreasing.

One can then define a distortion divergence as follows:

$$D_g(Q\|P) = \inf\{\eta \mid Q(A) \leq g_\eta(P(A)) \text{ for all } A \in \mathcal{F}\}. \tag{4.1}$$

If the set over which the infimum is taken in (4.1) is empty, which happens in particular when Q is not absolutely continuous with respect to P, then, in line with a standard convention, the value of the divergence is taken to be ∞. The set in (4.1) can be written as $\bigcap_{A\in\mathcal{F}}\{\eta \mid Q(A) \leq g_\eta(P(A))\}$; by the assumed continuity of the mapping $\eta \mapsto g_\eta(u)$ for fixed u, this shows that the set is closed, so that the infimum (if finite) is achieved and we might replace "inf" by "min" in the definition. In view of Prop. 3.1, the definition above can equivalently be written (for $Q \ll P$) as

$$D_g(Q\|P) = \min\{\eta \mid \mathrm{ROC}_{P,Q}(u) \leq g_\eta(u) \text{ for all } u \in [0,1]\}. \tag{4.2}$$

The following proposition verifies the basic property of divergences.

Proposition 4.2. *We have $D_g(Q\|P) \geq 0$ for all Q, and $D_g(Q\|P) = 0$ if and only if $P = Q$.*

Proof. The nonnegativity of the distortion divergence is immediate from the conditions in Assumption 4.1 on the parametrization of families of distortion functions. If $D_g(Q\|P) = 0$, then (because the infimum is attained) $Q(A) \leq P(A)$ for all $A \in \mathcal{F}$. Applying this inequality to an event A and its complement shows that in fact we have $Q(A) = P(A)$ for all $A \in \mathcal{F}$. Conversely, if $Q = P$, then $\mathrm{ROC}_{P,Q}(u) = u$ for all $u \in [0,1]$, so that the minimum in (4.1) is achieved in $\eta = 0$. □

For a given family of distortion functions g_η and a given number $r \geq 0$, one can define the divergence ball:

$$B_g^r = \{Q \mid D_g(Q\|P) \leq r\}. \tag{4.3}$$

Divergences are often used to identify points in a convex set which are "closest" (in the sense of the divergence) to a given point outside the set. For this purpose, the following proposition is important.

Proposition 4.3. *Distortion divergence balls are convex.*

Proof. A probability measure Q belongs to the ball B_g^r if and only if $Q(A) \leq g_r(P(A))$ for all $A \in \mathcal{F}$. If this property holds for two measures Q_1 and Q_2, then it certainly holds as well for any convex combination of Q_1 and Q_2. □

Basic concepts relating to risk measures can be expressed in terms of distortion divergence balls. For a given bounded random variable X, one can define

$$B_g^r(X) = \{E^Q[X] \mid Q \in B_g^r(P, Q)\}.$$

In other words, when X is considered as a functional acting on the space of probability measures that are absolutely continuous with respect to P, $B_g^r(X)$ is the image of the divergence ball B_g^r under this functional. It follows from the convexity of B_g^r that $B_g^r(X)$ is an interval. The condition for "acceptability" of the risk X in the sense of Artzner et al. (1999) can be rephrased as the requirement that this interval should be contained in $[0, \infty)$. The performance measures proposed in Cherny and Madan (2009) can be constructed as the maximal value of r for which a given position remains acceptable, i.e., the largest value of r such that $B_g^r(X) \subset [0, \infty)$.

The distortion divergence is not symmetric in general. The relation between the divergence from P to Q and the divergence from Q to P is given by the following proposition, which implies that symmetry of the divergence does hold if the underlying family of distortion functions is symmetric under reversion.

Proposition 4.4. *Suppose that the family $(g_\eta)_{\eta \geq 0}$ consists of distortion functions that are strictly increasing, and let the reverse family $(\hat{g}_\eta)_{\eta \geq 0}$ be defined by (2.7). For equivalent measures P and Q, the following relation holds:*

$$D_g(Q \| P) = D_{\hat{g}}(P \| Q). \tag{4.4}$$

Proof. For brevity of notation, write $f(u) := \mathrm{ROC}_{P,Q}(u)$ and $\hat{f}(u) := \mathrm{ROC}_{Q,P}(u)$. By Prop. 2.8, the function f is strictly increasing, and, for given $\eta \geq 0$, the condition $\hat{f}(u) \leq \hat{g}_\eta(u)$ for all $u \in [0, 1]$ is equivalent to the condition $f^{-1}(v) \geq g_\eta^{-1}(v)$ for all $v \in [0, 1]$. Because the mapping g_η is a bijection from $[0, 1]$ to itself, this in turn is equivalent to the condition $f^{-1}(g_\eta(u)) \geq u$ for all $u \in [0, 1]$, which finally is the same as requiring $f(u) \leq g_\eta(u)$ for all $u \in [0, 1]$. Therefore, we have

$$\min\{\eta \mid \hat{f}(u) \leq \hat{g}_\eta(u) \text{ for all } u \in [0, 1]\} = \min\{\eta \mid f(u) \leq g_\eta(u) \text{ for all } u \in [0, 1]\}.$$

This proves (4.4). □

The *total variation distance* between two probability measures P and Q on the same measurable space $(\Omega, \mathcal{F})$ is defined by (see for instance Denuit et al. (2005))

$$\delta(P, Q) = \sup\big\{ |Q(A) - P(A)| \;\big|\; A \in \mathcal{F}\big\}.$$

The following proposition provides an upper bound for the total variation distance in terms of the distortion divergence based on a given family of distortion functions.

Proposition 4.5. *Consider a family of distortion functions $(g_\eta)_{\eta\ge 0}$ satisfying Assumption 4.1. For $\eta \ge 0$, define*

$$m_g(\eta) = \max_{0\le u\le 1} g_\eta(u) - u. \tag{4.5}$$

Then, for probability measures P and Q with $Q \ll P$, we have

$$\delta(P,Q) \le m_g\big(D_g(Q\|P)\big). \tag{4.6}$$

Proof. Write $\eta := D_g(Q\|P)$ for brevity. We know that $Q(A) \le g_\eta(P(A))$ for all events A. From this it follows that $Q(A) - P(A) \le m_g(\eta)$ for all A. Applying the same inequality to the complement of A instead of A, we find that also $P(A) - Q(A) \le m_g(\eta)$ for all A. Therefore, the claim follows. $\square$

Example 4.6. In the case of the family of distortion functions defined by the Wang transform, the function defined in (4.5) is given by

$$m_W(\eta) = \Phi(\tfrac{1}{2}\eta) - \Phi(-\tfrac{1}{2}\eta) = 2\Phi(\tfrac{1}{2}\eta) - 1. \tag{4.7}$$

Therefore, we have the inequality

$$\delta(P,Q) \le 2\Phi\big(\tfrac{1}{2}D_W(Q\|P)\big) - 1. \tag{4.8}$$

Remark 4.7. Suppose that the distortion functions g_η are strictly increasing, and let $(\hat{g}_\eta)_{\eta\ge 0}$ denote the reverse family as defined in (2.7). From the interpretation of $\hat{g}$ as the reflection of g with respect to the cross diagonal in the unit square, it follows that $m_{\hat{g}} = m_g$. This implies that the upper bound that is obtained by reversing the roles of P and Q in (4.6) is the same as the one that is obtained by replacing the family g by its reverse:

$$m_g\big(D_g(P\|Q)\big) = m_g\big(D_{\hat{g}}(Q\|P)\big) = m_{\hat{g}}\big(D_{\hat{g}}(Q\|P)\big).$$

If the family g is not symmetric, then this upper bound is in general different from the one in (4.6).

Remark 4.8. The total variation distance can itself be constructed as a distortion divergence if the condition of continuity for distortion functions is dropped. In a parametrization with a parameter a that takes values between 0 and 1, the corresponding family of distortion functions is given by $g_a(0) = 0$ and $g_a(u) = \min(u + a, 1)$ for $0 < u \le 1$. For purposes of risk management, distortion functions that are not continuous at 0 would normally be considered too conservative, since the use of such distortions implies that alternative worlds are taken into account in which adverse events can receive a fixed positive minimal probability even when they are extremely unlikely under the reference (statistical) measure.

For some families of distortion risk measures, the distortion divergence can be written in a more explicit form, by making use of the ROC curve. Consider for instance the Wang transform $g_\eta(u) = \Phi(\Phi^{-1}(u) + \eta)$. The condition on the parameter η appearing in (4.2) can be written as

$$\mathrm{ROC}_{P,Q}(u) \le \Phi\big(\Phi^{-1}(u) + \eta\big).$$

This can in turn be rewritten as

$$\eta \geq \Phi^{-1}\big(\mathrm{ROC}_{P,Q}(u)\big) - \Phi^{-1}(u).$$

It follows that the distortion divergence based on the Wang transform can be expressed as

$$D_{\mathrm{W}}(Q\|P) = \sup_{0<u<1} \big[\Phi^{-1}\big(\mathrm{ROC}_{P,Q}(u)\big) - \Phi^{-1}(u)\big]. \tag{4.9}$$

In a similar way, one obtains expressions for divergences with respect to the PH family and with respect to the NV family of Example 3.9 and its reverse, as parametrized in Table 1, (3.8), and (3.9):

$$D_{\mathrm{PH}}(Q\|P) = \sup_{0<u<1} \frac{\log u}{\log \mathrm{ROC}_{P,Q}(u)} - 1 \tag{4.10}$$

$$D_{\mathrm{NV}}(Q\|P) = \sup_{0<u<1} \log \frac{\Phi^{-1}\big(\frac{1}{2}u\big)}{\Phi^{-1}\big(\frac{1}{2}\mathrm{ROC}_{P,Q}(u)\big)} \tag{4.11}$$

$$D_{\mathrm{RNV}}(Q\|P) = \sup_{0<u<1} \log \frac{\Phi^{-1}\big(\frac{1}{2}+\frac{1}{2}\mathrm{ROC}_{P,Q}(u)\big)}{\Phi^{-1}\big(\frac{1}{2}+\frac{1}{2}u\big)}. \tag{4.12}$$

In the case of CVaR, we arrive similarly at

$$D_{\mathrm{CVaR}}(Q\|P) = \sup_{0<u\leq 1} \frac{\mathrm{ROC}_{P,Q}(u)}{u} - 1.$$

This implies, by Prop. 2.2, that

$$D_{\mathrm{CVaR}}(Q\|P) = \operatorname{ess\,sup}_P \frac{dQ}{dP} - 1. \tag{4.13}$$

Up to monotone reparametrizaton, the distortion divergence induced by the CVaR family therefore coincides with the *separation divergence* introduced by Aldous and Diaconis (1987).

The following proposition gives an expression for the distortion divergence in the case of finitely discrete distributions. This proposition can serve as a basis for numerical methods.

Proposition 4.9. *Assume that the sample space* Ω *is finite, and that the outcomes are ordered according to (3.10). The following expression holds for the distortion divergence induced by a family* $(g_\eta)_{\eta\geq 0}$ *of distortion functions satisfying Assumption 4.1:*

$$D_g(Q\|P) = \min\{\eta \mid \textstyle\sum_{j=1}^{i} q_j \leq g_\eta\big(\sum_{j=1}^{i} p_j\big) \text{ for all } i = 1,\dots,n\}. \tag{4.14}$$

Proof. By (4.2), $D_g(Q,P)$ is the smallest value of η for which g_η is a majorant of $\mathrm{ROC}_{P,Q}$. Prop. 3.13 indicates that $\mathrm{ROC}_{P,Q}$ is the piecewise linear function interpolating the points (u_i, v_i) with $u_i := P(A_i)$ and $v_i := Q(A_i)$ $(i = 0,\dots,n)$. Because, for all η, g_η is concave, the inequality $\mathrm{ROC}_{P,Q}(u) \leq g_\eta(u)$ holds for all $u \in [0,1]$ if and only if $v_i \leq g_\eta(u_i)$ for $i = 0,\dots,n$. This proves the claim. □

The computation time needed for verification of the condition in (4.14) is linear in the number of elements of the sample space, as opposed to direct calculation on the basis of the definition (4.1) which would take exponential time. To compute the minimum in (4.14) in cases in which an analytic expression is not available, one might use for instance a bisection method.

Example 4.10. Let P and Q be two measures defined on $\Omega = \{0, 1\}$, and write $p := P(1)$, $q := Q(1)$. From the proposition above, one finds for instance

$$D_{\mathrm{W}}(Q\|P) = \left|\Phi^{-1}(q) - \Phi^{-1}(p)\right|$$

and

$$D_{\mathrm{PH}}(Q\|P) = \max\Big(\frac{\log p}{\log q}, \frac{\log(1-p)}{\log(1-q)}\Big) - 1.$$

5 Conclusions

The purpose of this note has been to highlight the connection between, on the one hand, distortion risk measures, and, on the other hand, optimal receiver operating characteristics associated to binary hypothesis testing problems. The role of the ROC curve with respect to distortion risk measures is threefold: first, as a way to generate families of distortion functions; second, to provide a statistical interpretation of distortion functions; and third, to facilitate computations.

Several families of distortion functions that have been proposed in the literature turn out to be related to well known hypothesis testing problems. The fact that one may create new families of distortion functions by consideration of specific families of hypothesis testing problems has been illustrated by making use of the problem of testing $N(0, 1)$ versus $N(0, \sigma^2)$. The connection between distortion functions and ROC curves leads to a statistical interpretation of distortion functions, which supports the transfer of a notion of deviation from the statistical domain to the risk management domain. In this way, researchers and practitioners may judge the degree of conservatism in choosing a particular parameter value in a family of distortion functions, even though there is no intrinsic relation between the risks to which the risk measure is applied and the distributions considered in the testing problem. ROC curves can also be put to good use in characterizing and computing the distortion divergence that is associated to a given one-parameter family of distortion functions.

A numerical method has been provided to find the distortion divergence between two given discrete measures. The computational time required is linear in the size of the sample space. The method can be applied to continuous measures after discretization. A more challenging computational task would be to compute the distortion divergence projection from a given probability measure onto a given set of probability measures, for instance defined by moment constraints; that is, to find a point (uniqueness may not always be guaranteed) in the given set that is closest to the given measure in the sense of the divergence. On the basis of characterizations as in (4.9), one is led to problems of minimax type.

References

David Aldous and Persi Diaconis. Strong uniform times and finite random walks. *Advances in Applied Mathematics*, 8:69–97, 1987.

Barry C. Arnold, Christopher A. Robertson, Patrick L. Brockett, and Boo-Yau Shu. Generating ordered families of Lorenz curves by strongly unimodal distributions. *Journal of Business & Economic Statistics*, 5:305–308, 1987.

Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

Hirbod Assa. On optimal reinsurance policy with distortion risk measures and premiums. *Insurance: Mathematics and Economics*, 61:70–75, 2015.

Anne Balter and Antoon Pelsser. Quantifying ambiguity bounds through hypothetical statistical testing. Working paper, 2017. `ssrn.com/abstract=2613843`.

Bank for International Settlements. Minimum capital requirements for market risk. `www.bis.org/bcbs/publ/d352.pdf`, January 2016.

Michèle Basseville. Divergence measures for statistical data processing—an annotated bibliography. *Signal Processing*, 93:621–633, 2013.

Jaume Belles-Sampera, José M. Merigó, Montserrat Guillén, and Miguel Santolino. The connection between distortion risk measures and ordered weighted averaging operators. *Insurance: Mathematics and Economics*, 52:411–420, 2013.

Patrick Billingsley. *Probability and Measure* (2nd ed.). Wiley, New York, 1986.

Tim J. Boonen. Competitive equilibria with distortion risk measures. *ASTIN Bulletin*, 45:703–728, 2015.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004.

Lev M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7:200–217, 1967.

Alexander Cherny and Dilip Madan. New measures for performance evaluation. *Review of Financial Studies*, 22:2571–2606, 2009.

Gustave Choquet. Theory of capacities. *Annales de l'institut Fourier*, 5:131–295, 1954.

Dieter Denneberg. *Non-Additive Measure and Integral*. Kluwer, Dordrecht, 1994.

Michel Denuit, Jan Dhaene, Marc Goovaerts, and Rob Kaas. *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. Wiley, Chichester, 2005.

Edward J. Dudewicz and Satya Mishra. *Modern Mathematical Statistics*. Wiley, New York, 1988.

Daniel Ellsberg. Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, 75:643–669, 1961.

Hans Föllmer and Alexander Schied. *Stochastic Finance: An Introduction in Discrete Time* (3rd ed.). Walter de Gruyter, Berlin, 2011.

Itzhak Gilboa and David Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18:141–153, 1989.

Thomas Goll and Ludger Rüschendorf. Minimax and minimal distance martingale measures and their relationship to portfolio optimization. *Finance and Stochastics*, 5:557–581, 2001.

Marc Goovaerts, F.E. De Vylder, and Jean Haezendonck. *Insurance Premiums*. North-Holland, Amsterdam, 1984.

Lars Peter Hansen and Thomas J. Sargent. *Robustness*. Princeton University Press, 2008.

Werner Hürlimann. Distortion risk measures and economic capital. *North American Actuarial Journal*, 8:86–95, 2004.

Werner Hürlimann. On the lookback distortion risk measure: theory and applications. *Journal of Mathematical Sciences: Advances and Applications*, 30:21–47, 2014.

Erich L. Lehmann and Joseph P. Romano. *Testing Statistical Hypotheses* (3rd ed.). Springer, New York, 2005.

Krzysztof Postek, Dick den Hertog, and Bertrand Melenberg. Computationally tractable counterparts of distributionally robust constraints on risk measures. *SIAM Review*, 58:603–650, 2016.

R. Mark Reesor and Donald L. McLeish. Risk, entropy, and the transformation of distributions. *North American Actuarial Journal*, 7(2):128–144, 2003.

R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1997.

David Schmeidler. Integral representation without additivity. *Proceedings of the American Mathematical Society*, 97:255–261, 1986.

Ekaterina N. Sereda, Efim M. Bronshtein, Svetozar T. Rachev, Frank J. Fabozzi, Wei Sun, and Stoyan V. Stoyanov. Distortion risk measures in portfolio optimization. In John B. Guerard Jr., editor, *Handbook of Portfolio Construction. Contemporary Applications of Markowitz Techniques*, pages 649–673. Springer, New York, 2010.

Andreas Tsanakas. Dynamic capital allocation with distortion risk measures. *Insurance: Mathematics and Economics*, 35:223–243, 2004.

Hideatsu Tsukahara. One-parameter families of distortion risk measures. *Mathematical Finance*, 19:691–705, 2009.

Abraham Wald. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16:117–186, 1945.

Shaun S. Wang. Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics*, 17:43–54, 1995.

Shaun S. Wang. Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, 26:71–92, 1996.

Shaun S. Wang. A class of distortion operators for pricing financial and insurance risks. *Journal of Risk and Insurance*, 67:15–36, 2000.

Shaun S. Wang, Virginia R. Young, and Harry H. Panjer. Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, 21:173–183, 1997.

Julia Lynn Wirch and Mary R. Hardy. A synthesis of risk measures for capital adequacy. *Insurance: Mathematics and Economics*, 25:337–347, 1999.

Virginia R. Young. Optimal insurance under Wang's premium principle. *Insurance: Mathematics and Economics*, 25:109–122, 1999.