# Probability metrics applied to problems in portfolio theory


Stoyan V. Stoyanov

FinAnalytica, Inc., USA

e-mail: stoyan.stoyanov@finanalytica.com

Svetlozar T. Rachev *

University of Karlsruhe, Germany and

University of California Santa Barbara, USA

e-mail: rachev@pstat.ucsb.edu

Frank J. Fabozzi

Yale University, School of Management

e-mail: frank.fabozzi@yale.edu

**Contact person:**
Prof. Svetlozar T. Rachev
Chair of Econometrics, Statistics
and Mathematical Finance
School of Economics and Business Engineering
University of Karlsruhe
Kollegium am Schloss, Bau II, 20.12, R210
Postfach 6980, D-76128, Karlsruhe, Germany



*Prof Rachev gratefully acknowledges research support by grants from Division of Mathematical, Life and Physical Sciences, College of Letters and Science, University of California, Santa Barbara, the Deutschen Forschungsgemeinschaft and the Deutscher Akademischer Austausch Dienst.

# Probability metrics applied to problems in portfolio theory

**Abstract**

In the paper, we consider the application of the theory of probability metrics in several areas in the field of finance. First, we argue that specially structured probability metrics can be used to quantify stochastic dominance relations. Second, the methods of the theory of probability metrics can be used to arrive at a general axiomatic treatment of dispersion measures and probability metrics can be used to describe continuity of risk measures. Finally, the methods of probability metrics theory are applied to the benchmark-tracking problem significantly extending the problem setting.

**Key words:** probability metrics, stochastic dominance, dispersion measure, deviation measure, risk measure, benchmark-tracking

# 1 Introduction

The development of the theory of probability metrics started with the investigation of problems related to limit theorems in probability theory. The limit theorems have a very important place in probability theory, statistics, and all their applications. A well-known example is the celebrated Central Limit Theorem (CLT), the Generalized CLT, the max-stable CLT, functional limit theorems, etc. In general, the applicability of the limit theorems stems from the fact that the limit law can be regarded as an approximation to the stochastic model considered and, therefore, can be accepted as an approximate substitute. The central question arising is how large an error we make by adopting the approximate model. This question can be investigated by studying the distance between the limit law and the stochastic model and whether it is, for example, the sum or maxima of independent and identically distributed (i.i.d.) random variables makes no difference as far as the universal principle is concerned.

Generally, the theory of probability metrics studies the problem of measuring distances between random quantities. On one hand, it provides the fundamental principles for building probability metrics — the means of measuring such distances. On the other, it studies the relationships between various classes of probability metrics. The second realm of study concerns problems which require a particular metric while the basic results can be obtained in terms of other metrics. In such cases, the metrics relationship is of primary importance.

Certainly, the problem of measuring distances is not limited to random quantities only. In its basic form, it originated in different fields of mathematics. Nevertheless, the theory of probability metrics was developed due to the need for metrics with specific properties. Their choice is very often dictated by the stochastic model under consideration and to a large extent determines the success of the investigation. Rachev (1991) provides more details on the methods of the theory of probability metrics and its numerous applications in both theoretical and more practical problems.

In this paper, our goal is to study the application of probability metrics in the field of financial economics and more specifically within the field of portfolio theory.[1] There are many problems which can be generalized by using probability metrics or extended by applying the methods of the theory. We start with a brief introduction into the theory of probability metrics. The axiomatic construction is described and interpretations of various metrics is

[1]Modern portfolio theory was first formulated by Markowitz (1952). In 1990 he was awarded the Nobel prize in economic sciences for this contribution.

given from a financial economics viewpoint. The first topic in financial economics that we discuss is the stochastic dominance theory which arises from expected utility theory. Expected utility theory is a fundamental approach for describing how choices under uncertainty are made. It is very basic not only for the field of finance but for microeconomic theory. The second financial economics topic is risk and dispersion measures. We generalize the axiomatic treatment of dispersion measures by probability metrics and quasi-metrics. Finally, we consider the benchmark-tracking problem in portfolio theory and its extension to relative deviation metrics which are constructed according to the methods of probability metrics theory.

# 2 Probability metrics

Generally speaking, a functional which measures the distance between random quantities is called a *probability metric*. These random quantities can be of a very general nature. For instance, with financial economics in view, they can be random variables (r.v.s), such as daily equity returns or daily exchange rate movements, or stochastic processes, such as the price evolution of a commodity in a given period, or much more complex objects such as the daily movement of the shape of the yield curve. We limit the discussion to one-dimensional r.v.s only. Rachev (1991) provides a more general treatment.

Probability metrics are defined axiomatically. Denote by $\mathfrak{X} := \mathfrak{X}(\mathbb{R})$ the set of all r.v.s on a given probability space $(\Omega, \mathfrak{A}, P)$ taking values in $(\mathbb{R}, \mathcal{B}_1)$ where $\mathcal{B}_1$ denotes the Borel $\sigma$-algebra of Borel subsets of $\mathbb{R}$, and by $\mathcal{L}\mathfrak{X}_2$ the space of all joint distributions $\text{Pr}_{X,Y}$ generated by the pairs $X, Y \in \mathfrak{X}$. Probability metrics are denoted by $\mu$ and are defined on the space of all joint distributions $\mathcal{L}\mathfrak{X}_2$, $\mu(X,Y) := \mu(\text{Pr}_{X,Y})$. The axiomatic construction is based on a number of properties which we list below. The formal definition is given afterwards.

Consider the following properties.

ID. $\mu(X,Y) \geq 0$ and $\mu(X,Y) = 0$, if and only if $X \sim Y$

$\widetilde{\text{ID}}$. $\mu(X,Y) \geq 0$ and $\mu(X,Y) = 0$, if $X \sim Y$

These two properties are called the *identity properties*. The notation $X \sim Y$ denotes that $X$ is equivalent to $Y$. The meaning of *equivalence* depends on the type of metrics. If the equivalence is in almost sure sense, then the metrics are called *compound*. If $\sim$ means equality of distribution, then the metrics are called *simple*. Finally, if $\sim$ stands for equality of some characteristics of

$X$ and $Y$, then the metrics are called *primary*. The axiom $\widetilde{\text{ID}}$ is weaker than ID.

The next axiom is called the *symmetry axiom*. It makes sense in the general context of calculating distances between elements of a space,

SYM. $\mu(X,Y) = \mu(Y,X)$

The third axiom is the *triangle inequality*,

TI. $\mu(X,Y) \leq \mu(X,Z) + \mu(Z,Y)$ for any $X, Y, Z$

The triangle inequality is important because it guarantees, together with ID, that $\mu$ is continuous in any of the two arguments,

$$|\mu(X,Y) - \mu(X,Z)| \leq \mu(Z,Y).$$

The triangle inequality can be relaxed to the more general form called *triangle inequality with parameter* $K$,

$\widetilde{\text{TI}}$. $\mu(X,Y) \leq K(\mu(X,Z) + \mu(Z,Y))$ for any $X, Y, Z$ and $K \geq 1$.

Notice that the traditional version TI appears when $K = 1$. Furthermore, the three pairs of r.v.s in $\widetilde{\text{TI}}$ should be chosen in such a way that there exists a consistent three-dimensional random vector $(X, Y, Z)$ and the three pairs are its two-dimensional projections.

The formal definition is given below.

**Definition 1.** *A mapping* $\mu : \mathcal{L}\mathfrak{X}_2 \to [0, \infty]$ *is said to be*

- *a probability metric if ID, SYM and TI hold,*
- *a probability semimetric if $\widetilde{ID}$, SYM, TI hold*
- *a probability distance with parameter $K_\mu$ if ID, SYM, and $\widetilde{TI}$ hold*
- *a probability semidistance with parameter $K_\mu$ if $\widetilde{ID}$, SYM, and $\widetilde{TI}$ hold*

## 2.1 Examples of probability distances

The difference between probability semi-metrics and probability semi-distances is in the relaxation of the triangle inequality. Probability semi-distances can be constructed from probability semi-metrics by means of an additional function $H(x) : [0, \infty) \to [0, \infty)$ which is non-decreasing and continuous and satisfies the following condition

$$K_H := \sup_{t>0} \frac{H(2t)}{H(t)} < \infty \tag{2.1}$$

which is known as *Orlicz's condition*. There is a general result which states that if $\rho$ is a metric function, then $H(\rho)$ is a semi-metric function and satisfies the triangle inequality with parameter $K = K_H$. We denote all functions satisfying the properties above and Orlicz's condition (2.1) by $\mathcal{H}$.

In this section, we provide examples of probability distances. We also provide interpretation of the formulae assuming that the random variables describe financial quantities.

### 2.1.1 Primary distances

Common examples of primary metrics include,

1. *The engineer's metric*

$$\mathbf{EN}(X, Y) := |EX - EY|$$

where $X$ and $Y$ are r.v.s with finite mathematical expectation, $EX < \infty$ and $EY < \infty$.

2. *The absolute moments metric*

$$\mathbf{MOM}_p(X, Y) := |m^p(X) - m^p(Y)|,\ p \geq 1$$

where $m^p(X) = (E|X|^p)^{1/p}$ and $X$ and $Y$ are r.v.s with finite moments, $E|X|^p < \infty$ and $E|Y|^p < \infty$, $p \geq 1$.

### 2.1.2 Simple distances

Common examples of simple metrics and distances are stated below.

1. *The Kolmogorov metric*

$$\boldsymbol{\rho}(X, Y) := \sup_{x \in \mathbb{R}} |F_X(x) - F_Y(x)| \tag{2.2}$$

where $F_X(x)$ is the distribution function of $X$ and $F_Y(x)$ is the distribution function of $Y$. The Kolmogorov metric is also called the *uniform* metric. It is applied in the CLT in probability theory.

If the r.v.s $X$ and $Y$ describe the return distribution of two common stocks, then the Kolmogorov metric has the following interpretation. The distribution function $F_X(x)$ is by definition the probability that $X$ loses more than a level $x$, $F_X(x) = P(X \leq x)$. Similarly, $F_Y(x)$ is the probability that $Y$ loses more than $x$. Therefore, the Kolmogorov distance $\boldsymbol{\rho}(X,Y)$ is the maximum deviation between the two probabilities that can be attained by varying the loss level $x$. If $\boldsymbol{\rho}(X,Y) = 0$, then the probabilities that $X$ and $Y$ lose more than a loss level $x$ coincide for all loss levels.

Usually, the loss level $x$, for which the maximum deviation is attained, is close to the mean of the return distribution, i.e. the mean return. Thus, the Kolmogorov metric is completely insensitive to the tails of the distribution which describe the probabilities of extreme events — extreme returns or extreme losses.

2. *The Lévy metric*

$$\mathbf{L}(X,Y) := \inf_{\varepsilon>0}\{F_X(x-\varepsilon)-\varepsilon \leq F_Y(x) \leq F_X(x+\varepsilon)+\varepsilon,\ \forall x \in \mathbb{R}\} \tag{2.3}$$

The Lévy metric is difficult to calculate in practice. It has important theoretic application in probability theory as it metrizes the weak convergence.

The Kolmogorov metric and the Lévy metric can be regarded as metrics on the space of distribution functions because $\boldsymbol{\rho}(X,Y) = 0$ and $\mathbf{L}(X,Y) = 0$ imply coincidence of the distribution functions $F_X$ and $F_Y$.

The Lévy metric can be viewed as measuring the closeness between the graphs of the distribution functions while the Kolmogorov metric is a uniform metric between the distribution functions. The general relationship between the two is

$$\mathbf{L}(X,Y) \leq \boldsymbol{\rho}(X,Y) \tag{2.4}$$

For example, suppose that $X$ is a r.v. describing the return distribution of a portfolio of stocks and $Y$ is a deterministic benchmark with a return of 2.5% ($Y = 2.5\%$). (The deterministic benchmark in this case could be either the cost of funding over a specified time period or a target return requirement to satisfy a liability such as a guaranteed investment contract.) Assume also that the portfolio return has a normal distribution with mean equal to 2.5% and a volatility $\sigma$. Since the expected portfolio return is exactly equal to the deterministic benchmark, the Kolmogorov distance between them is always equal to 1/2 irrespective of how small the volatility is,

$$\boldsymbol{\rho}(X, 2.5\%) = 1/2, \quad \forall\, \sigma > 0.$$

Thus, if we rebalance the portfolio and reduce its volatility, the Kolmogorov metric will not register any change in the distance between the portfolio return and the deterministic benchmark. In contrast to the Kolmogorov metric, the Lévy metric will indicate that the rebalanced portfolio is closer to the benchmark.

3. *The Kantorovich metric*

$$\boldsymbol{\kappa}(X, Y) := \int_{\mathbb{R}} |F_X(x) - F_Y(x)| dx. \tag{2.5}$$

where $X$ and $Y$ are r.v.s with finite mathematical expectation, $EX < \infty$ and $EY < \infty$.

The Kantorovich metric can be interpreted along the lines of the Kolmogorov metric. Suppose that $X$ and $Y$ are r.v.s describing the return distribution of two common stocks. Then, as we explained, $F_X(x)$ and $F_Y(x)$ are the probabilities that $X$ and $Y$, respectively, lose more than the level $x$. The Kantorovich metric sums the absolute deviation between the two probabilities for all possible values of the loss level $x$. Thus, the Kantorovich metric provides aggregate information about the deviations between the two probabilities.

In contrast to the Kolmogorov metric, the Kantorovich metric is sensitive to the differences in the probabilities corresponding to extreme profits and losses but to a small degree. This is because the difference $|F_X(x) - F_Y(x)|$ converges to zero as the loss level $(x)$ increases or decreases and, therefore, the contribution of the terms corresponding to extreme events to the total sum is small. As a result, the differences in the tail behavior of $X$ and $Y$ will be reflected in $\boldsymbol{\kappa}(X, Y)$ but only to a small extent.

4. *The Kantorovich distance*

$$\ell_H(X,Y) := \int_0^1 H(|F_X^{-1}(x) - F_Y^{-1}(x)|)dx, \quad H \in \mathcal{H} \tag{2.6}$$

where the r.v.s $X$ and $Y$ have finite mathematical expectation, $E|X| < \infty$, $E|Y| < \infty$. If we choose $H(t) = t^p,\ p \geq 1$, then $(\ell_H(X,Y))^{1/p}$ turns into the $L_p$ metric between inverse distribution functions, $\ell_p(X,Y)$, defined as

$$\ell_p(X,Y) := \left(\int_0^1 |F_X^{-1}(t) - F_Y^{-1}(t)|^p dt\right)^{1/\min(1,1/p)}, \ p > 0. \tag{2.7}$$

Under this slight extension, the limit case $p \to 0$ appears to be the total variation metric $\boldsymbol{\sigma}(X,Y)$

$$\ell_0(X,Y) = \boldsymbol{\sigma}(X,Y) := \sup_{\text{all events } A} |P(X \in A) - P(Y \in A)|. \tag{2.8}$$

The other limit case provides a relation to the uniform metric between inverse distribution functions $\mathbf{W}(X,Y)$,

$$\ell_\infty(X,Y) = \mathbf{W}(X,Y) := \sup_{0<t<1} |F_X^{-1}(t) - F_Y^{-1}(t)| \tag{2.9}$$

The uniform metric $\mathbf{W}(X,Y)$ has the following interpretation in finance. Suppose that $X$ and $Y$ describe the return distribution of two common stocks. Then the quantity $-F_X^{-1}(t)$ is known as the *value-at-risk* (VaR) of common stock $X$ at confidence level $(1-t)100\%$. It is used as a risk measure and represents a loss threshold such that losing more than it happens with probability $t$. The probability $t$ is also called the *tail probability* because the VaR is usually calculated for high confidence levels, e.g. 95%, 99%, and the corresponding loss thresholds are in the tail of the distribution.

Therefore, the difference $F_X^{-1}(t) - F_Y^{-1}(t)$ is nothing but the difference between the VaRs of $X$ and $Y$ at confidence level $(1-t)100\%$. Thus, the probability metric $\mathbf{W}(X,Y)$ is the maximal difference in absolute value between the VaRs of $X$ and $Y$ when the confidence level is varied. Usually, the maximal difference is attained for values of $t$ close to zero or one which corresponds to VaR levels close to the maximum loss or profit of the return distribution. As a result, the probability metric $\mathbf{W}(X,Y)$ is entirely centered on the extreme profits or losses.

5. *The Birnbaum-Orlicz average distance*

$$\boldsymbol{\theta}_H(X,Y) := \int_{\mathbb{R}} H(|F_X(x) - F_Y(x)|)dx, \quad H \in \mathcal{H} \tag{2.10}$$

where the r.v.s $X$ and $Y$ have finite mathematical expectation, $E|X| < \infty$, $E|Y| < \infty$. If we choose $H(t) = t^p,\ p \geq 1$, then $(\boldsymbol{\theta}_H(X,Y))^{1/p}$ turns into the $L_p$ metric between distribution functions, $\boldsymbol{\theta}_p(X,Y)$

$$\boldsymbol{\theta}_p(X,Y) := \left(\int_{-\infty}^{\infty} |F_X(t) - F_Y(t)|^p dt\right)^{1/\min(1,1/p)}, \ p > 0. \tag{2.11}$$

At limit as $p \to 0$,

$$\boldsymbol{\theta}_0(X,Y) := \int_{-\infty}^{\infty} I\{t : F_X(t) \neq F_Y(t)\}dt \tag{2.12}$$

where the notation $I\{A\}$ stands for the indicator of the set $A$. That is, the simple metric $\boldsymbol{\theta}_0(X,Y)$ calculates the Lebesgue measure of the set $\{t : F_X(t) \neq F_Y(t)\}$.

If $p \to \infty$, then we obtain the Kolmogorov metric defined in (2.2), $\boldsymbol{\theta}_\infty(X,Y) = \boldsymbol{\rho}(X,Y)$.

### 2.1.3 Compound distances

Common examples of compound metrics are stated below.

1. *The p-average compound metric*

$$\mathcal{L}_p(X,Y) = (E|X-Y|^p)^{1/p}, \ p \geq 1 \tag{2.13}$$

where $X$ and $Y$ are r.v.s with finite moments, $E|X|^p < \infty$ and $E|Y|^p < \infty$, $p \geq 1$.

From the viewpoint of finance, we can recognize two widely used measures of deviation which belong to the family of the $p$-average compound metrics. If $p$ is equal to one, we obtain the mean absolute deviation between $X$ and $Y$,

$$\mathcal{L}_1(X,Y) = E|X-Y|.$$

Suppose that $X$ describes the returns of a stock portfolio and $Y$ describes the returns of a benchmark portfolio. Then the mean absolute deviation is a way to measure how closely the stock portfolio tracks the benchmark.

2. *The Ky Fan metric*

$$\mathbf{K}(X,Y) := \inf\{\varepsilon > 0 : P(|X-Y| > \varepsilon) < \varepsilon\} \tag{2.14}$$

where $X$ and $Y$ are real-valued r.v.s. The Ky Fan metric has an important application in the theory of probability as it metrizes convergence in probability of real-valued random variables.

Assume that $X$ is a random variable describing the return distribution of a portfolio of stocks and $Y$ describes the return distribution of a benchmark portfolio. The probability

$$P(|X-Y| > \varepsilon) = P\left(\{X < Y - \epsilon\} \bigcup \{X > Y + \epsilon\}\right)$$

concerns the event that either the portfolio will outperform the benchmark by $\epsilon$ (i.e., earn a return that exceeds the return on the benchmark by $\epsilon$) or it will underperform the benchmark by $\epsilon$ (i.e., earn a return that is less than the benchmark by $\epsilon$). Therefore, the quantity $2\epsilon$ can be interpreted as the width of a performance band. The probability $1 - P(|X-Y| > \varepsilon)$ is actually the probability that the portfolio stays within the performance band.

As the width of the performance band decreases, the probability $P(|X-Y| > \varepsilon)$ increases. The Ky Fan metric calculates the width of a performance band such that the probability of the event that the portfolio return is outside the performance band is smaller than half of it.

3. *The Birnbaum-Orlicz compound metric*

$$\mathbf{\Theta}_p(X,Y) = \left(\int_{-\infty}^{\infty} \tau^p(t;X,Y)dt\right)^{1/p}, \ p \geq 1 \tag{2.15}$$

where $\tau(t;X,Y) = P(X \leq t < Y) + P(Y \leq t < X)$.

The function $\tau(t; X, Y)$, which is the building block of the Birnbaum-Orlicz compound metric, can be interpreted in the following way. Suppose that $X$ and $Y$ describe the return distributions of two common stocks. The function argument, $t$, can be regarded as a performance divide. The term $P(X \leq t < Y)$ is the probability that $X$ underperforms $t$ and, simultaneously, $Y$ outperforms $t$. If $t$ is a very small number, then the probability $P(X \leq t < Y)$ will be close to zero because the stock $X$ will underperform it very rarely. If $t$ is a very large number, then $P(X \leq t < Y)$ will again be close to zero because stock $Y$ will rarely outperform it. A similar conclusion holds for the other term of $\tau(t; X, Y)$ as it only treats the random variables in the opposite way. Therefore, we can conclude that the function $\tau(t; X, Y)$ calculates the probabilities of the relative underperformance or outperformance of $X$ and $Y$, and has a maximum for moderate values of the performance divide $t$.

In the case of $p = 1$, we have the following relationship,

$$\boldsymbol{\Theta}_1(X, Y) = E|X - Y| = \mathcal{L}_1(X, Y).$$

## 2.2 Ideal probability metrics

The ideal probability metrics are probability metrics which satisfy two additional properties which make them uniquely positioned to study problems related to limit theorems in probability theory. The two additional properties are the homogeneity property and the regularity property.

The *homogeneity property* of order $r \in \mathbb{R}$ is

HO. $\mu(cX, cY) = |c|^r \mu(X, Y)$ for any $X, Y$ and constant $c \in \mathbb{R}$.

The homogeneity property has the following interpretation in portfolio theory. If $X$ and $Y$ are r.v.s describing the random return of two portfolios, then converting proportionally into cash, for example, 30% of the two portfolios results in returns scaled down to $0.3X$ and $0.3Y$. Since the returns of the two portfolios appear scaled by the same factor, it is reasonable to assume that the distance between the two scales down proportionally.

The *regularity property* is

RE. $\mu(X + Z, Y + Z) \leq \mu(Y, X)$ for any $X, Y$ and $Z$

and the *weak regularity property* is

WRE. $\mu(X+Z, Y+Z) \leq \mu(Y, X)$ for any $X, Y$ and $Z$ independent of $X$ and $Y$.

The regularity property has the following interpretation in portfolio theory. Suppose that $X$ and $Y$ are r.v.s describing the random values of two portfolios and $Z$ describes the random price of a common stock. Then buying one share of stock $Z$ per portfolio results in two new portfolios with random wealth $X+Z$ and $Y+Z$. Because of the common factor in the two new portfolios, we can expect that the distance between $X+Z$ and $Y+Z$ is smaller than the one between $X$ and $Y$.

The formal definition of ideal probability metrics follows below.

**Definition 2.** *A compound probability semidistance $\mu$ is said to be an ideal probability semidistance of order $r$ if it satisfies properties HO and RE. If the semidistance is simple, we replace RE with WRE.*

The conditions which need to be satisfied in order for the ideal metrics to be finite are given below. Suppose that the probability metric $\mu(X, Y)$ is a simple ideal metric of order $r$. The finiteness of $\mu(X, Y)$ guarantees equality of all moments up to order $r$,

$$\mu(X, Y) < \infty \qquad \Longrightarrow \qquad E(X^k - Y^k) = 0, \quad k = 1, 2, \ldots, n < r.$$

Conversely, if all moments $k = 1, 2, \ldots, n < r$ agree and, in addition to this, the absolute moments of order $r$ are finite, then metric $\mu(X, Y)$ is finite,

$$\begin{array}{l} EX^k = EY^k \\ E|X|^r < \infty \\ E|Y|^r < \infty \end{array} \qquad \Longrightarrow \qquad \mu(X, Y) < \infty$$

where $k = 1, 2, \ldots, n < r$.

The conditions which guarantee finiteness of the ideal metric $\mu$ are very important when investigating the problem of convergence in distribution of random variables in the context of the metric $\mu$.[2] Consider a sequence of r.v.s $X_1, X_2, \ldots, X_n, \ldots$ and a r.v. $X$ which satisfy the conditions,

$$EX_n^k = EX^k, \ \forall n, \ k = 1, 2, \ldots, n < r$$

and

[2]It is said that the metric $\mu$ *metrizes* the convergence in distribution if a sequence of random variables $X_1, \ldots, X_n, \ldots$ converges in distribution to the random variable $X$, if and only if $\mu(X_n, X) \to 0$ as $n \to \infty$.

$$E|X|^r < \infty,\ E|X_n|^r < \infty,\ \forall n.$$

For all known ideal metrics $\mu(X, Y)$ of order $r > 0$, given the above moment assumptions, the following holds: $\mu(X_n, X) \to 0$ if and only if $X_n$ converges to $X$ in distribution and the absolute moment of order $r$ converge,

$$\mu(X_n, X) \to 0 \quad \text{if and only if} \quad X_n \xrightarrow{d} X \quad \text{and} \quad E|X_n|^r \to E|X^r|.$$

This result has the following interpretation. Suppose that $X$ and $Y$ describe the returns of two portfolios. Choose an ideal metric $\mu$ of order $3 < r < 4$, for example. The convergence result above means that if $\mu(X, Y) \approx 0$, then both portfolios have very similar distribution functions and also they have very similar means, volatilities and skewness.

Note that, generally, the c.d.f.s of two portfolios being "close" to each other does not necessarily mean that their moments will be approximately the same. It is of crucial importance which metric is chosen to measure the distance between the distribution functions. The ideal metrics have this nice property that they guarantee convergence of certain moments. Rachev (1991) provides an extensive review of the properties of ideal metrics and their application.

## 2.3 Examples of ideal probability metrics

There are examples of both compound and simple ideal probability metrics. For instance, the p-average compound metric $\boldsymbol{\mathcal{L}}_p(X, Y)$ defined in (2.13) and the Birnbaum-Orlicz metric $\boldsymbol{\Theta}_p(X, Y)$ defined in (2.15) are ideal compound probability metrics of order one and $1/p$ respectively. In fact, almost all known examples of ideal probability metrics of order $r > 1$ are simple metrics.

Almost all of the simple metrics discussed in the previous section are ideal. The last three examples include metrics which have not been discussed in the previous section.

1. The $L_p$-metrics between distribution functions $\boldsymbol{\theta}_p(X, Y)$ defined in equation (2.11) is an ideal probability metric of order $1/p$, $p \geq 1$.

2. The Kolmogorov metric $\boldsymbol{\rho}(X, Y)$ defined in equation (2.2) is an ideal metric of order 0. This can also be inferred from the relationship $\boldsymbol{\rho}(X, Y) = \boldsymbol{\theta}_\infty(X, Y)$.

3. The $L_p$-metrics between inverse distribution functions $\ell_p(X, Y)$ defined in equation (2.7) is an ideal metric of order 1.

4. The Kantorovich metric $\boldsymbol{\kappa}(X, Y)$ defined in equation (2.5) is an ideal metric of order 1. This can also be inferred from the relationship $\boldsymbol{\kappa}(X, Y) = \ell_1(X, Y)$.

5. The total variation metric $\boldsymbol{\sigma}(X, Y)$ defined in equation (2.8) is an ideal probability metric of order 0.

6. The uniform metric between inverse c.d.f.s $\mathbf{W}(X, Y)$ defined in equation (2.9) is an ideal metric of order 1.

7. *The Zolotarev ideal metric*

   The general form of the Zolotarev ideal metric is

$$\boldsymbol{\zeta}_s(X, Y) = \int_{-\infty}^{\infty} |F_{s,X}(x) - F_{s,Y}(x)|\, dx \tag{2.16}$$

   where $s = 1, 2, \dots$ and

$$F_{s,X}(x) = \int_{-\infty}^{x} \frac{(x-t)^{s-1}}{(s-1)!} dF_X(t) \tag{2.17}$$

   The Zolotarev metric $\boldsymbol{\zeta}_s(X, Y)$ is ideal of order $r = s$, see Zolotarev (1997).

8. *The Rachev metric*

   The general form of the Rachev metric is

$$\boldsymbol{\zeta}_{s,p,\alpha}(X, Y) = \left( \int_{-\infty}^{\infty} |F_{s,X}(x) - F_{s,Y}(x)|^p\, |x|^{\alpha p'} dx \right)^{1/p'} \tag{2.18}$$

   where $p' = \max(1, p)$, $\alpha \geq 0$, $p \in [0, \infty]$, and $F_{s,X}(x)$ is defined in equation (2.17). If $\alpha = 0$, then the Rachev metric $\boldsymbol{\zeta}_{s,p,0}(X, Y)$ is ideal of order $r = (s-1)p/p' + 1/p'$.

   Note that $\boldsymbol{\zeta}_{s,p,\alpha}(X, Y)$ can be represented in terms of lower partial moments,

$$\boldsymbol{\zeta}_{s,p,\alpha}(X, Y) = \frac{1}{(s-1)!} \left( \int_{-\infty}^{\infty} \left| E(t-X)_+^s - E(t-X)_+^s \right|^p |t|^{\alpha p'} dt \right)^{1/p'}.$$

9. *The Kolmogorov-Rachev metrics*

The Kolmogorov-Rachev metrics arise from other ideal metrics by a process known as *smoothing*. Suppose the metric $\mu$ is ideal of order $0 \leq r \leq 1$. Consider the metric defined as

$$\mu_s(X,Y) = \sup_{h \in \mathbb{R}} |h|^s \mu(X + hZ, X + hZ) \tag{2.19}$$

where $Z$ is independent of $X$ and $Y$ and is a symmetric random variable $Z \stackrel{d}{=} -Z$. The metric $\mu_s(X,Y)$ defined in this way is ideal of order $r = s$. Note that while (2.19) always defines an ideal metric of order $s$, this does not mean that the metric is finite. The finiteness of $\mu_s$ should be studied for every choice of the metric $\mu$.

The Kolmogorov-Rachev metrics are applied in estimating the convergence rate in the Generalized CLT and other limit theorems. Rachev and Rüschendorf (1998) and Rachev (1991) provide more background and further details on the application in limit theorems.

## 2.4 Minimal metrics

The minimal metrics have an important place in the theory of probability metrics. Denote by $\mu$ a given compound metric. The functional $\hat{\mu}$ defined by the equality

$$\hat{\mu}(X,Y) := \inf\{\mu(\widetilde{X}, \widetilde{Y}) : \widetilde{X} \stackrel{d}{=} X,\ \widetilde{Y} \stackrel{d}{=} Y\} \tag{2.20}$$

is said to be the minimal metric with respect to $\mu$.[3] The minimization preserves the essential triangle inequality with parameter $K_{\hat{\mu}} = K_\mu$ and also the identity property assumed for $\mu$.

Many of the well-known simple metrics arise as minimal metrics with respect to some compound metric. For example,

$$\begin{aligned} \ell_p(X,Y) &= \hat{\boldsymbol{\mathcal{L}}}_p(X,Y) \\ \boldsymbol{\theta}_p(X,Y) &= \hat{\boldsymbol{\Theta}}_p(X,Y). \end{aligned}$$

The Kolmogorov metric (2.2) can be represented as a special case of the simple metric $\boldsymbol{\theta}_p$,

[3]Rachev (1991) provides a mathematical proof that the functional defined by equation (2.20) is indeed a probability metric.

$$\boldsymbol{\rho}(X,Y) = \boldsymbol{\theta}_\infty(X,Y)$$

and, therefore, it also arises as a minimal metric

$$\boldsymbol{\rho}(X,Y) = \hat{\boldsymbol{\Theta}}_\infty(X,Y).$$

Not all simple metrics arise as minimal metrics. A compound metric such that its minimal metric is equivalent to a given simple metric is called *protominimal* with respect to the given simple metric. For instance, $\boldsymbol{\Theta}_1(X,Y)$ is protominimal to the Kantorovich metric $\boldsymbol{\kappa}(X,Y)$. As we noted, not all simple metrics have protominimal ones and, also, some simple metrics have several protominimal ones, see Rachev (1991) for further theory.

# 3 Stochastic orders and probability metrics

In this section, we illustrate an application of probability metrics in the theory of stochastic orders. In the field of finance, the theory of stochastic orders is closely related to the expected utility theory which describes how choices under uncertainty are made. The expected utility theory was introduced in von Neumann and Morgenstern (1944). According to it, investor's perferences are described in terms of an investor's *utility function*. If no uncertainty is present, the utility function can be interpreted as a mapping between the available alternatives and real numbers indicating the "relative happiness" the investor gains from a particular alternative. If an individual prefers good "A" to good "B", then the utility of "A" is higher than the utility of "B". Thus, the utility function characterizes individual's preferences. Von Neumann and Morgenstern showed that if there is uncertainty, then it is the *expected utility* which characterizes the preferences. The expected utility of an uncertain prospect, often called a *lottery*, is defined as the probability weighted average of the utilities of the simple outcomes.

Denote by $F_X(x)$ and $F_X(x)$ the c.d.f.s of two uncertain prospects $X$ and $Y$. An investor with utility function $u(x)$ prefers $X$ to $Y$, or is indifferent between them, if and only if the expected utility of $X$ is not below the expected utility of $Y$,

$$X \succeq Y \qquad \Longleftrightarrow \qquad Eu(X) \geq Eu(Y)$$

where

$$Eu(X) = \int_{\mathbb{R}} u(x) dF_X(x).$$

The basic result of von Neumann-Morgenstern is that the preference order of the investor, which should satisfy certain technical conditions, is represented by expected utility in which the investor's utility function is unique up to a positive linear transform.

Some properties of the utility function are derived from common arguments valid for investors belonging to a certain category. For example, concerning certain prospects, all investors who prefer more to less are called *non-satiable* and have non-decreasing utility functions, all risk-averse investors have concave utility functions, all investors favoring positive to negative skewness have utility functions with non-negative third derivative. In fact, assuming certain behavior of the derivatives of $u(x)$, we obtain utility functions representing different classes of investors.

Suppose that there are two portfolios $X$ and $Y$, such that all investors from a given class do not prefer $Y$ to $X$. This means that the probability distributions of the two portfolios differ in a special way that, no matter the particular expression of the utility function, if an investor belongs to the given class, then $Y$ is not preferred by that investor. In this case, we say that portfolio $X$ dominates portfolio $Y$ with respect to the class of investors. Such a relation is often called a *stochastic dominance relation* or a *stochastic ordering*.

Stochastic dominance relations of different orders are defined by assuming certain properties for the derivatives of $u(x)$. Denote by $\mathcal{U}_n$ the set of all utility functions, the derivatives of which satisfy the inequalities $(-1)^{k+1}u^{(k)}(x) \geq 0,\ k = 1, 2, \ldots, n$ where $u^{(k)}(x)$ denotes the $k$-th derivative of $u(x)$. Thus, for each $n$, we have a set of utility functions which is a subset of $\mathcal{U}_{n-1}$,

$$\mathcal{U}_1 \subset \mathcal{U}_2 \subset \ldots \subset \mathcal{U}_n \subset \ldots$$

Imposing certain properties on the derivatives of $u(x)$ requires that we make more assumptions for the moments of the random variables we consider. We assume that the absolute moments $E|X|^k$ and $E|Y|^k$, $k = 1, \ldots, n$ of the random variables $X$ and $Y$ are finite.

**Definition 3.** *We say that the portfolio $X$ dominates the portfolio $Y$ in the sense of the $n$-th order stochastic dominance, $X \succeq_n Y$, if no investor with a utility function in the set $\mathcal{U}_n$ would prefer $Y$ to $X$,*

$$X \succeq_n Y \quad \textit{if} \quad Eu(X) \geq Eu(Y),\ \forall u(x) \in \mathcal{U}_n.$$

Particular examples include the first-order stochastic dominance (FSD) which concerns the class of non-satiable investors, the second-order stochastic dominance (SSD) which concerns the non-satiable, risk-averse investors and so on.

There is an equivalent way of describing the $n$-th order stochastic dominance in terms of the c.d.f.s of the ventures only. The condition is,

$$X \succeq_n Y \quad \Longleftrightarrow \quad F_X^{(n)}(x) \leq F_Y^{(n)}(x), \ \forall x \in \mathbb{R} \tag{3.21}$$

where $F_X^{(n)}(x)$ stands for the $n$-th integral of the c.d.f. of $X$ defined recursively as

$$F_X^{(n)}(x) = \int_{-\infty}^{x} F_X^{(n-1)}(t)dt.$$

In fact, an equivalent form of the condition in (3.21) can be derived,

$$X \succeq_n Y \quad \Longleftrightarrow \quad E(t-X)_+^{n-1} \leq E(t-Y)_+^{n-1}, \ \forall t \in \mathbb{R} \tag{3.22}$$

where $(t-x)_+^{n-1} = \max(t-x,0)^{n-1}$. This equivalent formulation clarifies why it is necessary to assume that all absolute moments until order $n$ are finite.

Since in the $n$-th order stochastic dominance we furnish the conditions on the utility function as $n$ increases, the following relation holds,

$$X \succeq_1 Y \quad \Longrightarrow \quad X \succeq_2 Y \quad \Longrightarrow \quad \ldots \quad \Longrightarrow \quad X \succeq_n Y.$$

Further on, it is possible to extend the $n$-th order stochastic dominance to the $\alpha$-order stochastic dominance in which $\alpha \geq 1$ is a real number and instead of the ordinary integrals of the c.d.f.s, fractional integrals are involved. Ortobelli et al. (2007) provide more information on extensions of stochastic dominance orderings and their relation to probability metrics and risk measures.

The conditions for stochastic dominance involving the distribution functions of the ventures $X$ and $Y$ represent a powerful method to determine if an entire class of investors would prefer any of the portfolios. For example, in order to verify if any non-satiable, risk-averse investor would not prefer $Y$ to $X$, we have to verify if condition (3.21) holds with $n = 2$. Note that a negative result does not necessarily mean that any such investor would actually prefer $Y$ or be indifferent between $X$ and $Y$. It may be the case that the inequality between the quantities in (3.21) is satisfied for some values of the argument, and for others, the converse inequality holds. Thus, only a part of the non-satiable, risk-averse investors may prefer $X$ to $Y$; it now depends on the particular investor we consider.

Suppose the verification confirms that either $X$ is preferred or the investors are indifferent between $X$ and $Y$, $X \succeq_2 Y$. This result is only qualitative, there are no indications whether $Y$ would be categorically disregarded by all investors in the class, or the differences between the two portfolios are very small. Similarly, if we know that no investors from the class prefer $Y$ to $Z$, $Z \succeq_2 Y$, then can we determine whether $Z$ is more strongly preferred to $Y$ than $X$ is?

The only way to approach this question is to add a quantitative element through a probability metric since only by means of a probability metric can we calculate distances between random quantities. For example, we can choose a probability metric $\mu$ and we can calculate the distances $\mu(X, Y)$ and $\mu(Z, Y)$. If $\mu(X, Y) < \mu(Z, Y)$, then the return distribution of $X$ is "closer" to the return distribution of $Y$ than are the return distributions of $Z$ and $Y$. On this ground, we can draw the conclusion that $Z$ is more strongly preferred to $Y$ than $X$ is, on condition that we know in advance the relations $X \succeq_2 Y$ and $Z \succeq_2 Y$.

However, not any probability metric appears suitable for this calculation. Suppose that $Y$ and $X$ are normally distributed r.v.s describing portfolio returns with equal means, $X \in N(a, \sigma_X^2)$ and $Y \in N(a, \sigma_Y^2)$, with $\sigma_X^2 < \sigma_Y^2$. $Z$ is a prospect yielding $a$ dollars with probability one. The c.d.f.s $F_X(x)$ and $F_Y(x)$ cross only once at $x = a$ and the $F_X(x)$ is below $F_Y(x)$ to the left of the crossing point. Therefore, no risk-averse investor would prefer $Y$ to $X$ and consequently $X \succeq_2 Y$. The prospect $Z$ provides a non-random return equal to the expected returns of $X$ and $Y$, $EX = EY = a$, and, in effect, any risk-averse investor would rather choose $Z$ from the three alternatives, $Z \succeq_2 X \succeq_2 Y$.

A probability metric with which we would like to quantify the SSD order should be able to indicate that, first, $\mu(X, Y) < \mu(Z, Y)$ because $Z$ is more strongly preferred to $Y$ and, second, $\mu(Z, X) < \mu(Z, Y)$ because $Y$ is more strongly rejected than $X$ with respect to $Z$. The assumptions in the example give us the information to order completely the three alternatives and that is why we are expecting the two inequalities should hold.

Let us choose the Kolmogorov metric defined in equation (2.2). Applying the definition to the distributions in the example, we obtain that $\rho(X, Z) = \rho(Y, Z) = 1/2$ and $\rho(X, Y) < 1/2$. As a result, the Kolmogorov metric is capable of showing that $Z$ is more strongly preferred relative to $Y$ but cannot show that $Y$ is more strongly rejected with respect to $Z$.

The example shows that there are probability metrics which are not appropriate to quantify a stochastic dominance order. The task of finding a suitable metric is not a simple one because the structure of the metric should be based on the conditions defining the dominance order. Inevitably, we cannot expect that one probability metric will appear suitable for all stochastic orders, rather, a probability metric may be best suited for a selected stochastic dominance relation.

Technically, we have to impose another condition in order for the problem of quantification to have a practical meaning. The probability metric calculating the distances between the ordered r.v.s should be bounded. If $\mu(X,Y) = \infty$ and $\mu(Z,Y) = \infty$, then we cannot compare the investors' preferences.

Concerning the FSD order, a suitable choice for a probability metric is the Kantorovich metric defined in (2.5). Note that the condition in (3.21) with $n = 1$ can be restated as $F_X(x) - F_Y(x) \leq 0, \ \forall x \in \mathbb{R}$. Thus, summing up all absolute differences gives an idea how "close" $X$ is to $Y$ which is a natural way of measuring the distance between $X$ and $Y$ with respect to the FSD order. The Kantorovich metric is finite as long as the random variables have finite means. This is a natural assumption for applications in the field of financial economics.

In the general case of the $n$-th order stochastic dominance, the condition in equation (3.22) is very similar to the Rachev ideal metric $\boldsymbol{\zeta}_{s,p,0}(X,Y)$ given in equation (2.18). There are additional assumptions that have to be made for the r.v.s $X$ and $Y$ ensuring that the Rachev ideal metric is finite. These assumptions are related to the equality of certain moments.

## 3.1 Return versus payoff

The lotteries in expected utility theory are usually interpreted as probability distributions of payoffs. As a consequence, the stochastic dominance theory is usually applied to random payoffs instead to returns.

On the other hand, modern portfolio theory, as well as other cornerstone theories, is developed for random log-returns. It is argued that the investment return is a more important characteristic than investment payoff when comparing opportunities. In effect, when searching for consistency between modern portfolio theory and stochastic dominance, a problem arises. Even though log-returns and payoffs are directly linked, it turns out that, generally, stochastic dominance relations concerning two log-return distributions are not equivalent to the corresponding stochastic dominance relations concerning their payoff distributions. In this section, we establish a link between the two types of stochastic dominance relations.

Suppose that investors' preference relations are defined on random venture payoffs. That is, the domain of the utility function $u(x)$ is the positive half-line which is interpreted as the collection of all possible outcomes in terms of dollars from a given venture. Assume that the payoff distribution is actually the price distribution $P_t$ of a financial asset at a future time $t$. In line with the von Neumann-Morgenstern theory, the expected utility of $P_t$ for an investor with utility function $u(x)$ is given by

$$Eu(P_t) = \int_0^\infty u(x) dF_{P_t}(x) \tag{3.23}$$

where $F_{P_t}(x) = P(P_t \leq x)$ is the c.d.f. of the random variable $P_t$. Furthermore, suppose that the price of the financial asset at the present time is $P_0$. The expected utility of the log-return distribution has the form,

$$Ev(r_t) = \int_{-\infty}^\infty v(y) dF_{r_t}(y) \tag{3.24}$$

where $v(y)$ is the utility function of the investor on the space of log-returns which is unique up to a positive linear transform. Note that $v(y)$ is defined on the entire real line as the log-return can be any real number. The next proposition establishes a link between the two utility functions.

**Proposition 1.** *The relationships between the utility function $u(x)$, $x \geq 0$, defined on the random payoff of an investment and the utility function $v(y)$, $y \in \mathbb{R}$, defined on the random return of the same investment is given by,*

$$v(y) = a.u(P_0 \exp(y)) + b, \quad a > 0 \tag{3.25}$$

*and*

$$u(x) = c.v(\log(x/P_0)) + d, \quad c > 0. \tag{3.26}$$

*Proof.* Consider the substitution $x = P_0 \exp(y)$ in equation (3.23). Under the new variable, the c.d.f. of $P_t$ changes to

$$F_{P_t}(P_0 \exp(y)) = P(P_t \leq P_0 \exp(y)) = P\left(\log \frac{P_t}{P_0} \leq y\right)$$

which is, in fact, the distribution function of the log-return of the financial asset $r_t = \log(P_t/P_0)$. The integration range changes from the positive half-line to the entire real line and equation (3.23) becomes

$$Eu(P_t) = \int_{-\infty}^\infty u(P_0 \exp(y)) dF_{r_t}(y). \tag{3.27}$$

Compare equations (3.27) and (3.24). From the uniqueness of the expected utility representation, it appears that (3.27) is the expected utility of the log-return distribution. Therefore, the utility function $v(y)$ can be computed by means of the utility function $u(x)$ and the representation is unique up to a positive linear transform. □

Note that the two utilities in equations (3.27) and (3.24) are identical (up to a positive linear transform) and this is not surprising. In our reasoning, the investor is one and the same. We only change the way we look at the venture, in terms of payoff or log-return, but the venture is also fixed. As a result, we cannot expect that the utility gained by the investor will fluctuate depending on the point of view.

Because of the relationship between the functions $u$ and $v$, properties imposed on the utility function $u$ may not transfer to the function $v$ and vice versa. Concerning the $n$-th order stochastic dominance, the next proposition establishes a useful relationship.

**Proposition 2.** *Suppose that the utility function $v(y)$ from equation (3.24) belongs to the set $\mathcal{U}_n$, i.e. it satisfies the conditions*

$$(-1)^{k+1}v^{(k)}(y) \geq 0, \quad k = 1, 2, \ldots, n$$

*where $v^{(k)}(y)$ denotes the $k$-th derivative of $v(y)$. The function $u(x)$ given by (3.26) also belongs to the set $\mathcal{U}_n$. Furthermore, suppose that $P_0^1 = P_0^2$ are the present values of two financial assets with random prices $P_t^1$ and $P_t^2$ at some future time $t$. Then the following implication holds for $n > 1$*

$$P_t^1 \succeq_n P_t^2 \qquad \Longrightarrow \qquad r_t^1 \succeq_n r_t^2$$

*where $r_t^1$ and $r_t^2$ are the log-returns for the corresponding period. If $n = 1$, then*

$$P_t^1 \succeq_1 P_t^2 \qquad \Longleftrightarrow \qquad r_t^1 \succeq_1 r_t^2.$$

*Proof.* Denote by

$$\widetilde{\mathcal{U}}_n = \{u = f(v), v \in \mathcal{U}_n\}$$

where the transformation $f$ is defined by (3.26). The first statement is verified directly by differentiation. Thus, we establish that $\widetilde{\mathcal{U}}_n \subseteq \mathcal{U}_n$. Since the inverse transformation defined by (3.25) does not preserve the corresponding derivatives properties for $n > 1$, we have a strict inclusion, $\widetilde{\mathcal{U}}_n \subset \mathcal{U}_n$ for $n > 1$. If $n = 1$, then $\widetilde{\mathcal{U}}_1 = \mathcal{U}_1$.

Suppose that $P_t^1 \succeq_n P_t^2$, $n > 1$. Then, according to the definition of the stochastic dominance relation, $Eu(P_t^1) \geq Eu(P_t^2)$, $\forall u \in \mathcal{U}_n$. As a consequence, $Eu(P_t^1) \geq Eu(P_t^2)$, $\forall u \in \widetilde{\mathcal{U}}_n$. From the definition of the class $\widetilde{\mathcal{U}}_n$, the uniqueness of the expected utility representation, and the assumption that $P_0^1 = P_0^2$, we deduce that $Ev(r_t^1) \geq Ev(r_t^2)$, $\forall v \in \mathcal{U}_n$ and, therefore, $r_t^1 \succeq_n r_t^2$, $n > 1$. The same reasoning and the fact that $\widetilde{\mathcal{U}}_1 = \mathcal{U}_1$ proves the final claim of the proposition. □

Note that the condition $P_0^1 = P_0^2$ is important. If the present values of the two financial assets are not the same, then such relationships may not exist.

# 4 Dispersion measures

In financial economics, measures of dispersion are used to characterize the uncertainty related to a given quantity such as the stock returns for example. Generally, dispersion measures can be constructed by means of different descriptive statistics. They calculate how observations in a dataset are distributed, whether there is high or low variability around the mean of the distribution. Examples include the standard deviation, the interquantile range, and the mean-absolute deviation. The central absolute moment of order $k$ is defined as

$$m_k = E|X - EX|^k$$

and an example of a dispersion measure based on it is

$$(m_k)^{1/k} = (E|X - EX|^k)^{1/k}.$$

The common properties of the dispersion measures can be synthesized into axioms. Rachev et al. (2007) provide the following set of general axioms. We denote the dispersion measure of a r.v. $X$ by $D(X)$.

D1. $D(X + C) \leq D(X)$ for all $X$ and constants $C \geq 0$.

D2. $D(0) = 0$ and $D(\lambda X) = \lambda D(X)$ for all $X$ and all $\lambda > 0$.

D3. $D(X) \geq 0$ for all $X$, with $D(X) > 0$ for non-constant $X$.

According to D1, adding a positive constant does not increase the dispersion of a r.v. According to D2 and D3, the dispersion measure $D$ is equal to zero only if the r.v. is a constant. This property is very natural for any measure of dispersion. For example, it holds for the standard deviation, MAD, and semi-standard deviation.

An example of a dispersion measure satisfying these properties is the *colog measure* defined by

$$\text{colog}(X) = E(X \log X) - E(X)E(\log X).$$

where $X$ is a positive random variable. The colog measure is sensitive to additive shifts and has applications in finance as it is consistent with the preference relations of risk-averse investors, see Rachev et al. (2007).

## 4.1 Dispersion measures and probability metrics

Suppose that $\mu$ is a compound probability metric. In this case, if $\mu(X,Y) = 0$, it follows that the two random variables are coincident in all states of the world. Therefore, the quantity $\mu(X,Y)$ can be interpreted as a measure of relative deviation between $X$ and $Y$. A positive distance, $\mu(X,Y) > 0$, means that the two variables fluctuate with respect to each other and zero distance, $\mu(X,Y) = 0$, implies that there is no deviation of any of them relative to the other.

This idea is closely related to the notion of dispersion but it is much more profound because we obtain the notion of dispersion measures as a special case by considering the distance between $X$ and the mean of $X$, $\mu(X,EX)$. In fact, the functional $\mu(X,EX)$ provides a very general notion of a dispersion measure as it arises as a special case from a probability metric which represents the only general way of measuring distances between random quantities.

## 4.2 Deviation measures

Rockafellar et al. (2006) provide an axiomatic description of convex dispersion measures called *deviation measures*. Besides the axioms of dispersion measures, the deviation measures satisfy the property

D4. $D(X+Y) \leq D(X) + D(Y)$ for all $X$ and $Y$.

and D1 is replaced by

$\widetilde{\text{D1}}$. $D(X+C) = D(X)$ for all $X$ and constants $C \in \mathbb{R}$.

As a consequence of axiom $\widetilde{\text{D1}}$, the deviation measure is influenced only by the difference $X - EX$. If $X = EX$ in all states of the world, then the deviation measure is a constant and, therefore, it is equal to zero because of the positivity axiom. Conversely, if $D(X) = 0$, then $X = EX$ in all states of the world. Properties D2 and D4 establish the convexity of $D(X)$.

Apparently not all deviation measures are symmetric; that is, it is possible to have $D(X) \neq D(-X)$ if the random variable $X$ is not symmetric.

Nevertheless, symmetric deviation measures can easily be constructed. The quantity $\tilde{D}(X)$ is a symmetric deviation measure if we define it as

$$\tilde{D}(X) := \frac{1}{2}(D(X) + D(-X)),$$

where $D(X)$ is an arbitrary deviation measure.

## 4.3 Deviation measures and probability quasi-metrics

One of the axioms defining probability semidistances is the symmetry axiom SYM. In applications in financial economics, the symmetry axiom is not important and we can omit it. Thus, we extend the treatment of the defining axioms of probability semidistances in the same way as it is done in the field of functional analysis. In case the symmetry axiom, SYM, is omitted, then *quasi-* is added to the name.

**Definition 4.** *A mapping $\mu : \mathcal{L}\mathfrak{X}_2 \to [0, \infty]$ is said to be*

- *a probability quasi-metric if ID and TI hold,*
- *a probability quasi-semimetric if $\widetilde{ID}$ and TI hold,*
- *a probability quasi-distance if ID and $\widetilde{TI}$ hold,*
- *a probability quasi-semidistance if $\widetilde{ID}$ and $\widetilde{TI}$ hold.*

Note that by removing the symmetry axiom we obtain a larger class in which semimetrics appear as symmetric quasi-semimetrics.

In this section, we demonstrate that the deviation measures arise from probability quasi-metrics equipped with two additional properties — *translation invariance* and *positive homogeneity.* A probability quasi-metric is called *translation invariant* and *positively homogeneous* if it satisfies the following two properties

TINV. $\mu(X + Z, Y + Z) = \mu(Y, X)$ for any $X, Y, Z$.

PHO. $\mu(aX, aY) = a\mu(X, Y)$ for any $X, Y$ and $a > 0$.

**Proposition 3.** *The functional $\mu_D$ defined as*

$$\mu_D(X, Y) = D(X - Y) \tag{4.28}$$

*is a positively homogeneous, translation invariant probability quasi-semimetric if $D$ is a deviation measure. Furthermore, the functional $D_\mu$ defined as*

$$D_\mu(X) = \mu(X - EX, 0) \tag{4.29}$$

*is a deviation measure if $\mu$ is a a positively homogeneous, translation invariant probability quasi-metric.*

*Proof.* We start with the first statement in the proposition. We verify if $\mu_D$ defined in equation (4.28) satisfies the necessary properties.

$\widetilde{\text{ID}}$. $\mu_D(X,Y) \geq 0$ follows from the non-negativity of $D$, property D3. Further on, if $X = Y$ in almost sure sense, then $X - Y = 0$ in almost sure sense and $\mu_D(X,Y) = D(0) = 0$ from Property D2.

TI. Follows from property D4:

$$\begin{aligned}\mu(X,Y) &= D(X-Y) = D(X - Z + (Z - Y)) \\ &\leq D(X-Z) + D(Z-Y) = \mu(X,Z) + \mu(Z,Y)\end{aligned}$$

TINV. A direct consequence of the definition in (4.28).

PHO. Follows from property D2.

We continue with the second statement in the proposition. We verify if $D_\mu$ defined in equation (4.28) satisfies the necessary properties.

$\widetilde{\text{D1}}$. A direct consequence of the definition in (4.29).

D2. Follows from ID and PHO. $D_\mu(0) = \mu(0,0) = 0$ and

$$D_\mu(\lambda X) = \lambda\mu(X - EX, 0) = \lambda D_\mu(X)$$

D3. Follows because $\mu$ is a probability metric. If $D_\mu(X) = 0$, then $X - EX$ is equal to zero almost surely which means that $X$ is a constant in all states of the world.

D4. Arises from TI and TINV.

$$\begin{aligned} D(X+Y) &= \mu(X - EX + Y - EY, 0) = \mu(X - EX, -Y + EY) \\ &\leq \mu(X - EX, 0) + \mu(0, -Y + EY) \\ &= \mu(X - EX, 0) + \mu(Y - EY, 0) \\ &= D(X) + D(Y) \end{aligned}$$

□

As a corollary from the proposition, all symmetric deviation measures arise from the translation invariant, positively homogeneous probability metrics.

Note that because of the properties of deviation measures, $\mu_D$ is a quasi-semimetric and cannot become a quasi-metric. This is because $D$ is not sensitive to additive shifts and this property is inherited by $\mu_D$,

$$\mu_D(X + a, Y + b) = \mu_D(X, Y),$$

where $a$ and $b$ are constants. In effect, $\mu_D(X, Y) = 0$ implies that the two random variables differ by a constant, $X = Y + c$, in all states of the world.

Due to the translation invariance property, equation (4.29) can be equivalently re-stated as

$$D_\mu(X) = \mu(X, EX). \tag{4.30}$$

In fact, as we remarked, equation (4.30) represents a very natural generic way of defining measures of dispersion. Starting from equation (4.30) and replacing the translation invariance property (TINV) by the weak regularity property (WRE) of ideal probability metrics, the sub-additivity property (D4) of $D_\mu(X)$ breaks down and a property similar to D1 holds instead of $\widetilde{\text{D1}}$,

$$D_\mu(X + C) = \mu(X + C, EX + C) \leq \mu(X, EX) = D_\mu(X)$$

for all constants $C$. In fact, this property is more general than D1 as it holds for arbitrary constants.

# 5 Risk measures

We have remarked that probability metrics provide the only way of measuring distances between random quantities. It turns out that a small distance between random quantities does not necessarily imply that selected characteristics of those quantities will be close to each other. If we want small distances measured by a probability metric to imply similar characteristics, the probability metric should be carefully chosen.

In finance, a risk measure $\rho$ is defined as the mapping $\rho : \mathfrak{X} \to \mathbb{R}$. It can be viewed as calculating a particular characteristic of a r.v. $X$. There are problems in finance in which the goal is to find a r.v. closest to another r.v. For instance, such is the benchmark tracking problem which is at the heart of passive portfolio construction strategies. Essentially, we are trying to construct a portfolio so as to track the performance a given benchmark. In some sense, this can be regarded as finding a portfolio return distribution which is closest to the return distribution of the benchmark. Usually, the distance is measured through the standard deviation of the difference $r_p - r_b$ where $r_p$ is the portfolio return and $r_b$ is the benchmark return.[4]

Suppose that we have found the portfolio tracking the benchmark most closely with respect to the tracking error. Generally, the risk of the portfolio is close to the risk of the benchmark only if we use the standard deviation as a risk measure because of the inequality,

$$|\sigma(r_p) - \sigma(r_b)| \leq \sigma(r_p - r_b).$$

The right part corresponds to the tracking error and, therefore, smaller tracking error results in $\sigma(r_p)$ being closer to $\sigma(r_b)$.

In order to guarantee that the small distance between the portfolio return distributions corresponds to similar risks, we have to find a suitable probability metric. Technically, for a given risk measure we need to find a probability metric with respect to which the risk measure is a continuous functional,

$$|\rho(X) - \rho(Y)| \leq \mu(X, Y),$$

where $\rho$ is the risk measure and $\mu$ stands for the probability metric. We continue with examples of how this can be done for the value-at-risk (VaR) and average value-at-risk (AVaR).

1. VaR

   The VaR at confidence level $(1 - \epsilon)100\%$, or tail probability $\epsilon$, is defined as the negative of the lower $\epsilon$-quantile of the return or payoff distribution,

$$VaR_\epsilon(X) = -\inf_x \{x | P(X \leq x) \geq \epsilon\} = -F_X^{-1}(\epsilon) \quad (5.31)$$

[4] In the parlance of portfolio management, this is quantity is referred to as the "active return".

where $\epsilon \in (0,1)$ and $F_X^{-1}(\epsilon)$ is the inverse of the distribution function of $X$.

Suppose that $X$ and $Y$ describe the return distributions of two portfolios. The absolute difference between the VaRs of the two portfolios at any tail probability can be bounded by,

$$\begin{aligned} |VaR_\epsilon(X) - VaR_\epsilon(Y)| &\leq \max_{p\in(0,1)} |VaR_p(X) - VaR_p(Y)| \\ &= \max_{p\in(0,1)} |F_Y^{-1}(p) - F_X^{-1}(p)| \\ &= \mathbf{W}(X,Y) \end{aligned}$$

where $\mathbf{W}(X,Y)$ is the uniform metric between inverse distribution functions defined in equation (2.9). If the distance between $X$ and $Y$ is small, as measured by the metric $\mathbf{W}(X,Y)$, then the VaR of $X$ is close to the VaR of $Y$ at any tail probability level $\epsilon$.

2. AVaR

The AVaR at tail probability $\epsilon$ is defined as the average of the VaRs which are larger than the VaR at tail probability $\epsilon$. Therefore, by construction, the AVaR is focused on the losses in the tail which are larger than the corresponding VaR level. The average of the VaRs is computed through the integral

$$AVaR_\epsilon(X) := \frac{1}{\epsilon}\int_0^\epsilon VaR_p(X)dp \tag{5.32}$$

where $VaR_p(X)$ is defined in equation (5.31).

Suppose that $X$ and $Y$ describe the return distributions of two portfolios. The absolute difference between the AVaRs of the two portfolios at any tail probability can be bounded by,

$$\begin{aligned} |AVaR_\epsilon(X) - AVaR_\epsilon(Y)| &\leq \frac{1}{\epsilon}\int_0^\epsilon |F_X^{-1}(p) - F_Y^{-1}(p)|dp \\ &\leq \int_0^1 |F_X^{-1}(p) - F_Y^{-1}(p)|dp \\ &= \kappa(X,Y) \end{aligned}$$

where $\kappa(X,Y)$ is the Kantorovich metric defined in equation (2.5). If the distance between $X$ and $Y$ is small, as measured by the metric

$\kappa(X, Y)$, then the AVaR of $X$ is close to the AVaR of $Y$ at any tail probability level $\epsilon$. Note that the quantity

$$\kappa_\epsilon(X, Y) = \frac{1}{\epsilon} \int_0^\epsilon |F_X^{-1}(p) - F_Y^{-1}(p)| dp$$

can also be used to bound the absolute difference between the AVaRs. It is a probability semi-metric giving the best possible upper bound on the absolute difference between the AVaRs.

# 6 Strategy replication

An important problem for fund managers is comparing the performance of their portfolios to a benchmark. The benchmark could be a market index or any other portfolio. In general, there are two types of strategies that managers follow: active or passive. An active portfolio strategy uses available information and forecasting techniques to seek a better performance than a portfolio that is simply diversified broadly. Essential to all active strategies are expectations about the factors that could influence the performance of an asset class. The goal of an active strategy is to outperform the benchmark after management fees by a given number of basis points. A passive portfolio strategy involves minimal expectational input and instead relies on diversification to match the performance of some benchmark. In effect, a passive strategy, commonly referred to as indexing, assumes that the marketplace will reflect all available information in the price paid for securities. There are various strategies for constructing a portfolio to replicate the index but the key in these strategies is designing a portfolio whose tracking error relative to the benchmark is as small as possible. Tracking error is the standard deviation of the difference between the return on the replicating portfolio and the return on the benchmark.

The benchmark tracking problem can be formulated as the optimization problem

$$\min_{w \in \mathcal{W}} \sigma(w'r - r^b)$$

where $w = (w_1, \dots, w_n)$ is a vector of portfolio weights, $\mathcal{W}$ is a set of admissible vectors $w$, $r = (r_1, \dots, r_n)$ is a vector of stocks returns, $r^b$ is the return of a benchmark portfolio, $w'r = \sum_{i=1}^n w_i r_i$ is the return of the portfolio in which $w_i$ is the weight of the $i$-th stock with return $r_i$, and $\sigma(X)$ stands for the standard deviation of the random variable $X$. The goal is to find a portfolio which is closest to the benchmark in a certain sense, in this case, the "closeness" is determined by the standard deviation. Each feasible vector of weights $w$ defines a portfolio with return $w'r$. Therefore, where appropriate, instead of $\mathcal{W}$ we use $\mathcal{X}$ to denote the feasible set of random variables $w'r$.

A serious disadvantage of the tracking error is that it penalizes in the same way the positive and the negative deviations from the mean excess return while our attitude towards them is asymmetric, see, among others, Szegö (2004) and the references therein. There is overwhelming evidence from the literature in the field of behavioral finance that people pay more attention to losses than to respective gains. This argument leads to the conclusion that a more realistic measure of "closeness" should be asymmetric.

The minimal tracking error problem can be restated in the more general form

$$\min_{w \in \mathcal{W}} \mu(w'r, r^b) \tag{6.33}$$

where $\mu(X, Y)$ is a measure of the deviation of $X$ relative to $Y$. Due to this interpretation, we regard $\mu$ as a functional which measures relative deviation and we call it a *relative deviation metric* or simply, r.d. metric.

In Stoyanov et al. (2007), it is argued that a reasonable assumption for the r.d. metrics is that they are positively homogeneous, regular quasi-semimetrics satisfying the additional property

$$\mu(X + c_1, Y + c_2) = \mu(X, Y) \text{ for all } X, Y \text{ and constants } c_1, c_2.$$

In fact, this property is always satisfied if we consider the functional $\mu$ on the sub-space of zero-mean random variables.

As a corollary, this property allows measuring the distance only between the centered portfolios returns because $\mu(X - EX, Y - EY) = \mu(X, Y)$. It may be argued that in practice the expected return of the portfolio is a very important characteristic and it seems that we are eliminating it from the problem. This is certainly not the case because this characteristic, as some others, can be incorporated into the constraint set $\mathcal{W}$ of problem (6.33). For example, a reasonable candidate for a constraint set of a long-only portfolio problem[5] is

$$\mathcal{W} = \{w : w'e = 1,\ w'Er \geq Er^b\}.$$

where $e = (1, \ldots, 1)$ and $w'Er \geq Er^b$ means that the optimal portfolio should have an expected return that is not below the benchmark.

[5] In portfolio management, a long-only portfolio is one in which only long positions in common stocks are allowed. A long position means the ownership of a stock. A short position means that the stock was sold short.

## 6.1 Examples of r.d. metrics

We distinguish between simple and compound quasi-semimetrics and the same distinction is valid for the r.d. metrics.

### 6.1.1 Compound metrics

We can illustrate how a probability metric can be modified so that it becomes an r.d. metric. Let us choose two classical examples of compound probability metrics — the average compound metric $\mathcal{L}_p(X,Y)$ defined in (2.13) and the Birnbaum-Orlicz compound metric $\mathbf{\Theta}_p(X,Y)$ defined in (2.15).

Consider, first, the average compound metric. It satisfies all necessary properties but it is symmetric, a property we would like to break. One possible way is to replace the absolute value by the max function. Thus we obtain the asymmetric version

$$\mathcal{L}_p^*(X,Y) = (E(\max(X-Y,0))^p)^{1/p},\ p \geq 1. \tag{6.34}$$

In Stoyanov et al. (2007) we show that $\mathcal{L}_p^*(X,Y)$ is an ideal quasi-semimetric; that is, using the max function instead of the absolute value breaks only the symmetry axiom SYM.

The intuition behind removing the absolute value and considering the max function is the following. In the setting of the benchmark-tracking problem, suppose that the r.v. $X$ stands for the return of the benchmark and $Y$ denotes the return of the portfolio. Minimizing $\mathcal{L}_p^*(X,Y)$, we actually decrease the average portfolio underperformance.

The same idea, but implemented in a different way, stays behind the asymmetric version of the Birnbaum-Orlicz metric

$$\mathbf{\Theta}_p^*(X,Y) = \left[\int_{-\infty}^{\infty} (\tau^*(t;X,Y))^p dt\right]^{1/p},\ p \geq 1 \tag{6.35}$$

where $\tau^*(t;X,Y) = P(Y \leq t < X)$. Stoyanov et al. (2007) show that (6.35) is an ideal quasi-semimetric. That is, considering only the first summand of the function $\tau(t;X,Y)$ from the Birnbaum-Orlics compound metric breaks the SYM axiom only.

Just as in the case of $\mathcal{L}_p^*(X,Y)$, suppose that the r.v. $Y$ represents the return of the portfolio and $X$ represents the benchmark return. Then, for a fixed value of the argument $t$, which we interpret as a threshold, the function $\tau^*$ calculates the probability of the event that the portfolio return is below the threshold $t$ and, simultaneously, the benchmark return is above the threshold. As a result, we can interpret $\mathbf{\Theta}_p^*(X,Y)$ as a measure of the probability that the portfolio loses more than the benchmark. Therefore, in the benchmark-tracking problem, by minimizing $\mathbf{\Theta}_p^*(X,Y)$, we are indirectly minimizing the probability of the portfolio losing more than the benchmark.

In order for (6.34) and (6.35) to become r.d. metrics, we consider them on the sub-space of zero-mean random variables.

### 6.1.2 Simple r.d. metrics

Simple r.d. metrics can be obtained through the minimization formula in equation (2.20). It is possible to show that, if $\mu$ is a functional satisfying properties ID or $\widetilde{\text{ID}}$, TI or $\widetilde{\text{TI}}$, then $\hat{\mu}$ also satisfies ID or $\widetilde{\text{ID}}$, TI or $\widetilde{\text{TI}}$. That is, omitting the symmetry property results only in asymmetry in the minimal functional $\hat{\mu}$. In addition, it is easy to check that if PHO holds for $\mu$, then the same property holds for $\hat{\mu}$ as well. These results, and one additional concerning convexity, are collected in the following proposition. Concerning the regularity property RE, there is a separate theorem which guarantees that if $\mu$ is regular, so is the minimal functional $\hat{\mu}$.

**Proposition 4.** *Suppose that $\mu$ is a positively homogeneous, compound quasi-semimetric. Then $\hat{\mu}$ defined in (2.20) is a positively homogeneous, simple quasi-semimetric. If $\mu$ satisfies the convexity property*

$$\mu(aX + (1-a)Y, Z) \leq a\mu(X,Z) + (1-a)\mu(Y,Z) \tag{6.36}$$

*for any $X, Y, Z$, then $\hat{\mu}$ satisfies*

$$\hat{\mu}(a\widetilde{X} + (1-a)\widetilde{Y}, Z) \leq a\hat{\mu}(X,Z) + (1-a)\hat{\mu}(Y,Z) \tag{6.37}$$

*where the pairs of r.v.s $(\widetilde{X},\widetilde{Z})$, $(\widetilde{Y},\widetilde{Z})$, $\widetilde{X} \stackrel{d}{=} X$, $\widetilde{Y} \stackrel{d}{=} Y$ and $\widetilde{Z} \stackrel{d}{=} Z$ are such that the minimum in (2.20) is attained.*

*Proof.* We prove only that the minimal metric satisfies the PHO property and (6.37). The remaining facts are proved in Rachev (1991), page 27. The PHO property is straightforward to check,

$$\begin{aligned}
\hat{\mu}(aX, aY) &= \inf\{\mu(\widetilde{X},\widetilde{Y}) : \widetilde{X} \stackrel{d}{=} aX, \widetilde{Y} \stackrel{d}{=} aY\} \\
&= \inf\{a^s\mu(\widetilde{X}/a,\widetilde{Y}/a) : \widetilde{X}/a \stackrel{d}{=} X, \widetilde{Y}/a \stackrel{d}{=} Y\} \\
&= a^s \inf\{\mu(\widetilde{X}/a,\widetilde{Y}/a) : \widetilde{X}/a \stackrel{d}{=} X, \widetilde{Y}/a \stackrel{d}{=} Y\} \\
&= a^s\hat{\mu}(X,Y).
\end{aligned}$$

Assume that the compound metric $\mu$ is convex in the sense of (6.36) in which $X$, $Y$, $Z$ are arbitrary r.v.s. We can always find pairs of r.v.s $(\widetilde{X},\widetilde{Z})$,

$(\widetilde{Y}, \widetilde{Z})$ such that $\widetilde{X} \stackrel{d}{=} X$, $\widetilde{Y} \stackrel{d}{=} Y$ and $\widetilde{Z} \stackrel{d}{=} Z$, and also $\hat{\mu}(X, Z) + \epsilon \geq \mu(\widetilde{X}, \widetilde{Z})$ and $\hat{\mu}(Y, Z) + \epsilon \geq \mu(\widetilde{Y}, \widetilde{Z})$, in which $\hat{\mu}$ denotes the minimal metric. In Rachev (1991), page 27, it is proved that the two bivariate laws can be consistently embedded in a triple $(\widetilde{X}, \ \widetilde{Y}, \ \widetilde{Z})$ so that the corresponding bivariate projections are the given pairs. Since (6.36) is true for any choice of three r.v.s, it will be true for the triple $(\widetilde{X}, \ \widetilde{Y}, \ \widetilde{Z})$,

$$\begin{aligned} \mu(a\widetilde{X} + (1-a)\widetilde{Y}, \widetilde{Z}) &\leq a\mu(\widetilde{X}, \widetilde{Z}) + (1-a)\mu(\widetilde{Y}, \widetilde{Z}), \ a \in (0,1) \\ &\leq a\hat{\mu}(X, Z) + (1-a)\hat{\mu}(Y, Z) + \epsilon \end{aligned} \tag{6.38}$$

Knowing the triple $(\widetilde{X}, \ \widetilde{Y}, \ \widetilde{Z})$, we can calculate the distribution of the convex combination $\widetilde{X}_a = a\widetilde{X} + (1-a)\widetilde{Y}$. In a similar vein, we can find a pair $(\bar{X}_a, \bar{Z})$ such that $\bar{X}_a \stackrel{d}{=} \widetilde{X}_a$, $\bar{Z} \stackrel{d}{=} \widetilde{Z} \stackrel{d}{=} Z$ and $\hat{\mu}(\widetilde{X}_a, Z) + \epsilon \geq \mu(\bar{X}_a, \bar{Z})$ which, applied to the left hand-side of (6.38), yields

$$\hat{\mu}(\widetilde{X}_a, Z) \leq \mu(a\widetilde{X} + (1-a)\widetilde{Y}, Z)$$

As a result, letting $\epsilon \to 0$ in (6.38), we obtain

$$\hat{\mu}(\widetilde{X}_a, Z) \leq a\hat{\mu}(X, Z) + (1-a)\hat{\mu}(Y, Z).$$

□

The convexity condition (6.36) is important for the optimal properties of the minimization problem in (6.33). The condition (6.36) ensures that the optimization problem is convex. Unfortunately, the convexity property breaks down for the minimal r.d. metric. The resulting property in (6.37) is weak to guarantee nice optimal properties of the minimization problem in (6.33).

Sometimes, it is possible to calculate explicitly the minimal functional. The Cambanis-Simons-Stout theorem provides explicit forms of the minimal functional with respect to a compound functional having the general form

$$\mu_\phi(X, Y) := E\phi(X, Y)$$

where $\phi(x, y)$ is a specific function called *quasi-antitone*, see Cambanis et al. (1976). The function $\phi(x, y)$ is called quasi-antitone if it satisfies the following property

$$\phi(x, y) + \phi(x', y') \leq \phi(x', y) + \phi(x, y') \tag{6.39}$$

for any $x' > x$ and $y' > y$. This property is related to how the function increases when its arguments increase. Also, the function $\phi$ should satisfy the technical condition that $\phi(x,x) = 0$. General examples of quasi-antitone functions include

a) $\phi(x,y) = f(x-y)$ where $f$ is a non-negative convex function in $\mathbb{R}$, for instance $\phi(x,y) = |x-y|^p$, $p \geq 1$.

b) $\phi(x,y) = -F(x,y)$ where $F(x,y)$ is the distribution function of a two dimensional random variable.

**Theorem 1.** *(Cambanis-Simons-Stout) Given $X, Y \in \mathfrak{X}$ with finite moments $E\phi(X,a) < \infty$ and $E\phi(Y,a) < \infty$, $a \in \mathbb{R}$ where $\phi(x,y)$ is a quasi-antitone function, then*

$$\widehat{\mu}_\phi(X,Y) = \int_0^1 \phi(F_X^{-1}(t), F_Y^{-1}(t))dt$$

*in which $F_X^{-1}(t) = \inf\{x : F_X(x) \geq t\}$ is the generalized inverse of the c.d.f. $F_X(x)$ and also $\widehat{\mu}_\phi(X,Y) = \mu_\phi(F_X^{-1}(U), F_Y^{-1}(U))$ where $U$ is a uniformly distributed r.v. on $(0,1)$.*

Applying the Cambanis-Simons-Stout theorem to the compound functional in equation (6.34), we obtain

$$l_p^*(X,Y) = \left[\int_0^1 (\max(F_X^{-1}(t) - F_Y^{-1}(t), 0))^p dt\right]^{1/p}, \ p \geq 1. \tag{6.40}$$

where $X$ and $Y$ are zero-mean random variables.

Besides the Cambanis-Simons-Stout theorem, there is another method of obtaining explicit forms of minimal and maximal functionals. This method is, essentially, direct application of the Fréchet-Hoeffding inequality between distribution functions,

$$\begin{aligned} \max(F_X(x) + F_Y(y) - 1, 0) \leq P(X \leq x, Y \leq y) \\ \leq \min(F_X(x), F_Y(y)). \end{aligned}$$

We show how this inequality is applied to the problem of finding the minimal r.d. metric of the Birnbaum-Orlicz quasi-semi-metric defined in (6.35) by taking advantage of the upper bound.

Consider the following representation of the $\tau^*$ function defined in (6.35),

$$\begin{aligned}\tau^*(t; X, Y) &= P(Y \leq t, X < t) \\ &= P(Y \leq t) - P(Y \leq t, X \leq t).\end{aligned}$$

Now by replacing the joint probability by the upper bound from the Frechet-Hoeffding inequality, we obtain

$$\begin{aligned}\tau^*(t; X, Y) &\geq F_Y(t) - \min(F_X(t), F_Y(t)) \\ &= \max(F_Y(t) - F_X(t), 0).\end{aligned}$$

Raising both sides of the above inequality to the power $p \geq 1$ and integrating over all values of $t$ does not change the inequality. In effect, we obtain

$$\left[\int_{-\infty}^{\infty} (\max(F_Y(t) - F_X(t), 0))^p dt\right]^{1/p} \leq \mathbf{\Theta}_p^*(X, Y)$$

which gives, essentially, the corresponding minimal r.d. metric,

$$\theta_p^*(X, Y) = \left[\int_{-\infty}^{\infty} (\max(F_Y(t) - F_X(t), 0))^p dt\right]^{1/p}, \; p \geq 1. \tag{6.41}$$

where $X$ and $Y$ are zero-mean random variables.

We have demonstrated that the Cambanis-Simons-Stout theorem and the Frechet-Hoeffding inequality can be employed to obtain explicitly the minimal functionals in equations (6.40) and (6.41),

$$\begin{aligned}l_p^*(X, Y) &= \hat{\mathcal{L}}_p^*(X, Y) \\ \boldsymbol{\theta}_p^*(X, Y) &= \hat{\mathbf{\Theta}}_p^*(X, Y)\end{aligned}$$

### 6.1.3 An example on the convexity of r.d. metrics

In this section, we provide an example illustrating that the convexity property (6.36) in Proposition 4 does not hold for the minimal r.d. metric. The example is based on the functional $\mu(X, Y) = \mathcal{L}_2(X, Y)$ defined in (2.13). It is a compound metric satisfying the condition in (6.36). We show that the minimal metric $\ell_2(X, Y) = \hat{\mathcal{L}}_2(X, Y)$ given in (2.7) does not satisfy the convexity property (6.36).

Suppose that $X \in N(0, \sigma_X^2)$, $Y \in N(0, \sigma_Y^2)$ and $Z \in N(0, \sigma_Z^2)$. Then we can calculate a closed-form expression for the minimal metric,

$$
\begin{aligned}
l_2(X,Y) &= \left(\int_0^1 (F_X^{-1}(t) - F_Y^{-1}(t))^2 dt\right)^{1/2} \\
&= \left(\int_0^1 (\sigma_X F^{-1}(t) - \sigma_Y F^{-1}(t))^2 dt\right)^{1/2} \\
&= \left(\int_0^1 \left((\sigma_X - \sigma_Y) F^{-1}(t)\right)^2 dt\right)^{1/2} \\
&= |\sigma_X - \sigma_Y| \left(\int_0^1 (F^{-1}(t))^2 dt\right)^{1/2} \\
&= |\sigma_X - \sigma_Y|
\end{aligned}
\tag{6.42}
$$

in which $F^{-1}$ is the inverse of the c.d.f. of the standard normal distribution.

In order to illustrate the convexity property, we have to calculate the distribution of the convex combination $\widetilde{X}_a = a\widetilde{X} + (1-a)\widetilde{Y}$, $0 \leq a \leq 1$. In it, the pair $(\widetilde{X}, \widetilde{Y})$ is a bivariate projection of the triple $(\widetilde{X}, \widetilde{Y}, \widetilde{Z})$ which is a three-dimensional vector of r.v.s having as two dimensional projections the pairs $(\widetilde{X}, \widetilde{Z})$ and $(\widetilde{Y}, \widetilde{Z})$ yielding the minimal metric. In the case of the $\mathcal{L}_2(X,Y)$ metric, these two bivariate projections can be computed explicitly from the Cambanis-Simons-Stout theorem,

$$
\begin{aligned}
(\widetilde{X}, \widetilde{Z}) &= (F_X^{-1}(U), F_Z^{-1}(U)) \\
(\widetilde{Y}, \widetilde{Z}) &= (F_Y^{-1}(U), F_Z^{-1}(U))
\end{aligned}
$$

where $U$ is a uniformly distributed r.v. on $(0,1)$. This result shows that the r.v.s are functionally dependent and $(\widetilde{X}, \widetilde{Y}) = (F_X^{-1}(U), F_Y^{-1}(U))$. This bivariate distribution corresponds to a bivariate Gaussian law with perfectly positively correlated components. As a result, the distribution of the convex combination is Gaussian, $\widetilde{X}_a \in N(0, (a\sigma_X + (1-a)\sigma_Y)^2)$. Therefore, by equation (6.42), the left hand-side of (6.37) equals,

$$
l_2(\widetilde{X}_a, Z) = |a\sigma_X + (1-a)\sigma_Y - \sigma_Z|.
$$

It remains to verify if the inequality in (6.37) holds. This is a straightforward calculation,

$$
\begin{aligned}
l_2(\widetilde{X}_a, Z) &= |a\sigma_X + (1-a)\sigma_Y - \sigma_Z| \\
&= |a(\sigma_X - \sigma_Z) + (1-a)(\sigma_Y - \sigma_Z)| \\
&\leq |a(\sigma_X - \sigma_Z)| + |(1-a)(\sigma_Y - \sigma_Z)| \\
&= a|\sigma_X - \sigma_Z| + (1-a)|\sigma_Y - \sigma_Z| \\
&= al_2(X, Z) + (1-a)l_2(Y, Z).
\end{aligned}
$$

Note that the bivariate distribution of $(\widetilde{X}, \widetilde{Y})$ is not the same as the bivariate law $(X, Y)$, which can be arbitrary, even though the marginals are the same. Therefore, the distribution of the convex combination $\widetilde{X}_a = a\widetilde{X} + (1-a)\widetilde{Y}$ is not the same as the distribution of $X_a = aX + (1-a)Y$ because of the different dependence. As a consequence, $\ell_2(\widetilde{X}_a, Z) \neq \ell_2(X_a, Z)$ which is the reason the convexity property (6.36) does not hold for the minimal functional. For example, if we assume that the bivariate law $(X, Y)$ has a zero-mean bivariate normal distribution with some covariance matrix, then $X_a \in N(0, \sigma_{aX+(1-a)Y})$. In this case,

$$
\ell_2(X_a, Z) = |\sigma_{aX+(1-a)Y} - \sigma_Z|,
$$

which may not be a convex function of $a$.

# 7 Conclusion

In this paper, we discuss the connections between the theory of probability metrics and the field of financial economics, particularly portfolio theory. We considered the theories of stochastic dominance, risk and dispersion measures, and benchmark-tracking problems and we found that the theory of probability metrics has appealing applications. Probability metrics can be used to quantify the dominance relations, they generalize the treatment of dispersion measures, and they offer a fundamental approach to generalizing the benchmark-tracking problem.

Even though in the paper we consider static problems, the generality of the suggested approach allows for extensions in a dynamic setting by studying probability metrics not in the space of random variables but in the space of random processes.

# References


Cambanis, S., G. Simons and W. Stout (1976), 'Inequalities for Ek(X,Y) when the marginals are fixed', *Z. Wahrsch. Verw. Geb.* **36**, 285–294.

Markowitz, H. M. (1952), 'Portfolio selection', *Journal of Finance* **7, (1)**, 77–91.

Ortobelli, S., S. T. Rachev, H. Shalit and F. J. Fabozzi (2007), 'Risk probability functionals and probability metrics applied to portfolio theory', *working paper* .

Rachev, S. and L. Rüschendorf (1998), *Mass transportation problems, V1, V2*, Springer-Verlag, NY.

Rachev, S. T. (1991), *Probability Metrics and the Stability of Stochastic Models*, Wiley, Chichester, U.K.

Rachev, S. T., S. Ortobelli, S. Stoyanov, F. J. Fabozzi and A. Biglova (2007), 'Desirable properties of an ideal risk measure in portfolio theory', *forthcoming in International Journal of Theoretical and Applied Finance* .

Rockafellar, R. T., S. Uryasev and M. Zabarankin (2006), 'Generalized deviations in risk analysis', *Finance and Stochastics* **10**, 51–74.

Stoyanov, S., S. Rachev, S. Ortobelli and F. Fabozzi (2007), 'Relative deviation metrics and the problem of strategy replication', *forthcoming in Journal of Banking and Finance* .

Szegö, G. (2004), *Risk measures for the* $21^{st}$ *century*, Wiley & Son Chichester.

von Neumann, J. and O. Morgenstern (1944), *Theory of games and economic behavior*, Princeton University Press, Princeton, N.J.

Zolotarev, V. M. (1997), *Modern theory of summation of random variable*, Brill Academic Publishers.