Lecture 3: Probability metrics

Prof. Dr. Svetlozar Rachev

Institute for Statistics and Mathematical Economics
University of Karlsruhe

Portfolio and Asset Liability Management

Summer Semester 2008

Copyright

The material is based on the text-book:

Svetlozar T. Rachev, Stoyan Stoyanov, and Frank J. Fabozzi
Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization: The Ideal Risk, Uncertainty, and Performance Measures
John Wiley, Finance, 2007

Prof. Svetlozar (Zari) 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
Tel. +49-721-608-7535, +49-721-608-2042(s)
Fax: +49-721-608-3811
http://www.statistik.uni-karslruhe.de

Introduction

- Theory of probability metrics came from the investigations related to limit theorems in probability theory.
- A well-known example is Central Limit Theorem (CLT) but there are many other limit theorems, such as the Generalized CLT, the max-stable CLT, functional limit theorems, etc.
- The limit law can be regarded as an approximation to the stochastic model considered and, therefore, can be accepted as an approximate substitute.
- 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, sum or maxima of i.i.d. random variables makes no difference as far as the universal principle is concerned.

Introduction

- The theory of probability metrics studies the problem of measuring distances between random quantities.
- First, it provides the fundamental principles for building probability metrics — the means of measuring such distances.
- Second, it studies the relationships between various classes of probability metrics.
- It also concerns problems which require a particular metric while the basic results can be obtained in terms of other metrics.

- No limitations in the theory of probability metrics on the nature of the random quantities makes its methods fundamental and appealing.
- It is more appropriate to refer to the random quantities as random elements: random variables, random vectors, random functions or random elements of general spaces.
- For instance, in the context of financial applications, we can study the distance between two random stocks prices, or between vectors of financial variables building portfolios, or between entire yield curves which are much more complicated objects.

$\Rightarrow$ The methods of the theory remain the same, no matter the nature of the random elements.

Measuring distances: the discrete case

So how can we measure the distance between two random quantities?

The important topics will be discussed such as:

- Examples of metrics defined on sets of characteristics of discrete distributions
- Examples of metrics based on the cumulative distribution function of discrete random variables
- Examples of metrics defined on the joint probability of discrete random variables
- Minimal and maximal distances

Sets of characteristics

Let us consider a pair of unfair dice and label the elements of the pair "die X" and "die Y".

"Die X" face	1	2	3	4	5	6
Probability, p_i	3/12	2/12	1/12	2/12	2/12	2/12
"Die Y" face	1	2	3	4	5	6
Probability, q_i	2/12	2/12	2/12	1/12	2/12	3/12

Table: The probabilities of the faces of "die X" and "die Y"

In the case of "die X", the probability of face 1 is higher than 1/6, which is the probability of a face of a fair die, and the probability of face 3 is less than 1/6. The probabilities of "die Y" have similar deviations from those of a fair die.

Measuring distances: the discrete case

- We can view the pair of dice as an example of two discrete random variables: X for “die X” and Y for “die Y”.
- The two discrete random variables have different distributions and, also, different characteristics, such as the mean and higher moments.
- Therefore, we can compare the two random variables in terms of the differences in some of their characteristics.
- For example, let us choose the mathematical expectation:

$$EX = \sum_{i=1}^{6} ip_i = 40/12 \quad \text{and} \quad EY = \sum_{i=1}^{6} iq_i = 44/12.$$

Measuring distances: the discrete case

- The distance between the two random variables, $\mu(X, Y)$, may be computed as the absolute difference between the corresponding mathematical expectations,

$$\mu(X, Y) = |EX - EY| = 4/12.$$

- The second moment can be calculated:

$$EX^2 = \sum_{i=1}^{6} i^2 p_i = 174/12 \quad \text{and} \quad EY^2 = \sum_{i=1}^{6} i^2 q_i = 202/12.$$

- If we add it to the mathematical expectation, for the distance we obtain

$$\mu(X, Y) = |EX - EY| + |EX^2 - EY^2| = 32/12.$$

- If we considered a pair of fair dice, these characteristics would coincide and we would obtain that the distance between the two random variables is zero.
- However, it is possible to obtain zero deviation between given characteristics in the case of unfair dice.
- Let us illustrate this with the variance of X and Y. The variance of a random variable Z, DZ, is defined as,

$$DZ = E(Z - EZ)^2.$$

or

$$DZ = E(Z - EZ)^2 = EZ^2 - (EZ)^2.$$

Measuring distances: the discrete case

- The variance of X equals

$$DX = EX^2 - (EX)^2 = \frac{174}{12} - \left(\frac{40}{12}\right)^2 = \frac{61}{18}$$

and the variance of Y equals

$$DY = EY^2 - (EY)^2 = \frac{202}{12} - \left(\frac{44}{12}\right)^2 = \frac{61}{18}.$$

- We obtain that $DX = DY$.

$\Rightarrow$ Thus, any attempts to measure the distance between the two random variables in terms of differences in variance will indicate zero distance even though “die X” is quite different from “die Y”.

Distribution functions

- By including more additional characteristics when measuring the distance between two random variables, we incorporate in $\mu(X, Y)$ more information from their distribution functions.
- How many characteristics we have to include, when X and Y have discrete distributions, so that we can be sure that the entire distribution function of X, $F_X(x) = P(X \leq x)$ agrees to the entire distribution of Y, $F_Y(x) = P(Y \leq x)$?
- Let us consider

$$\mu(X, Y) = \sum_{k=1}^{n} |EX^k - EY^k| \quad (1)$$

assuming that X and Y are the two dice considered above but this time we do not know the probabilities p_i and q_i, $i = 1, 6$.

Distribution functions

- How large should n be so that $\mu(X, Y) = 0$ guarantees that the distributions of X and Y agree completely?
  Since $\mu(X, Y) = 0$ is equivalent to

$$\left| \begin{array}{l} EX = EY \\ EX^2 = EY^2 \\ \dots \\ EX^n = EY^n \end{array} \right. \iff \left| \begin{array}{l} \sum_{i=1}^{6} i(p_i - q_i) = 0 \\ \sum_{i=1}^{6} i^2(p_i - q_i) = 0 \\ \dots \\ \sum_{i=1}^{6} i^n(p_i - q_i) = 0 \end{array} \right.$$

  then we need exactly 5 equations in order to guarantee that $P(X = i) = p_i = P(Y = i) = q_i, i = 1, 6$.
- Because there are 6 differences $p_i - q_i$ in the equations and we need 6 equations from the ones above plus the additional equation $\sum_{i=1}^{6}(p_i - q_i) = 0$ as all probabilities should sum up to one.

$\Rightarrow$ If X and Y are positive integers valued with k outcomes, then we need $k - 1$ equations in order to solve the linear system.

Distribution functions

$\Rightarrow$ If a given number of characteristics of two discrete random variables with finitely many outcomes agree, then their distribution functions agree completely.

- Then, instead of trying to figure out how many characteristics to include in a metric of a given type, is it possible to consider ways of measuring the distance between X and Y directly through their distribution function?
- If the distribution functions of two random variables coincide, then we have equal corresponding probabilities of any event and we can conclude that they have the same probabilistic properties.
- In the pair of dice example, all events are described by the set of all possible unions of the outcomes.

Distribution functions

- The distribution functions $F_X(x)$ and $F_Y(x)$ of “die X” and “die Y” are easy to calculate,

$$F_X(x) = \begin{cases} 0, & [x] < 1 \\ \sum_{i=1}^{[x]} p_i, & [x] \geq 1 \end{cases} = \begin{cases} 0, & x < 1 \\ 3/12, & 1 \leq x < 2 \\ 5/12, & 2 \leq x < 3 \\ 6/12, & 3 \leq x < 4 \\ 8/12, & 4 \leq x < 5 \\ 10/12, & 5 \leq x < 6 \\ 1, & x \geq 6 \end{cases} \quad (2)$$

$$F_Y(x) = \begin{cases} 0, & [x] < 1 \\ \sum_{i=1}^{[x]} q_i, & [x] \geq 1 \end{cases} = \begin{cases} 0, & x < 1 \\ 2/12, & 1 \leq x < 2 \\ 4/12, & 2 \leq x < 3 \\ 6/12, & 3 \leq x < 4 \\ 7/12, & 4 \leq x < 5 \\ 9/12, & 5 \leq x < 6 \\ 1, & x \geq 6 \end{cases} \quad (3)$$

where $[x]$ denotes the largest integer smaller than x.

Distribution functions

- One way to calculate the distance between two discrete cumulative distribution functions (c.d.f.s) $F_X(x)$ and $F_Y(x)$ is to *calculate the maximal absolute difference between them*,

$$\mu(X, Y) = \max_{x \in \mathbb{R}} |F_X(x) - F_Y(x)|. \quad (4)$$

- In the case of the two dice example, equation (4) can be readily computed, $\max_{x \in \mathbb{R}} |F_X(x) - F_Y(x)| = 1/12$. The maximum is attained at any $x \in [1, 3) \bigcup [4, 6)$.

Distribution functions

Another approach is to *compute the area closed between the graphs of the two functions*. If the area is zero, then due to the properties of the c.d.f.s we can conclude that the two functions coincide.

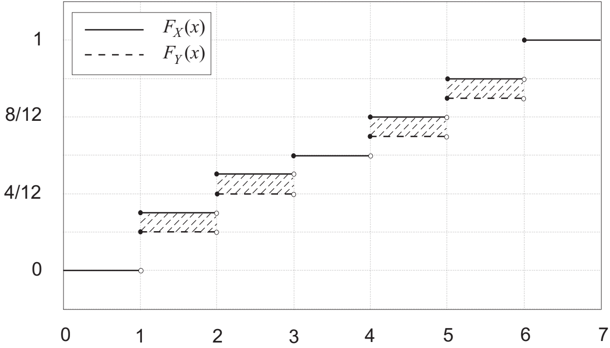

Figure: The plot shows the c.d.f.s of “die X” and “die Y”. The area closed between the graphs of the two c.d.f.s is shaded.

- The formula for the total area between the graphs of the two step functions is easy to arrive at,

$$\mu(X, Y) = \sum_{k=1}^{6} \left| \sum_{i=1}^{k} p_i - \sum_{i=1}^{k} q_i \right| . \tag{5}$$

- Using the probabilities given in the table before on the slide 7, we compute that the $\mu(X, Y) = 4/12$.

Distribution functions

- A similar approach can be adopted with respect to the quantile function of a random variable Z, or the inverse of the c.d.f.
- If the inverse c.d.f.s of two random variables coincide, then the distribution functions coincide. Then the distance between two random variables can be measured through the distance between the inverse of the c.d.f.s.
- The inverse $F_Z^{-1}(t)$ of the c.d.f. is defined as

$$F_Z^{-1}(t) = \inf\{x : F_Z(x) \geq t\}.$$

Distribution functions

- For example, the inverse c.d.f.s of (2) and (3) are

$$F_X^{-1}(t) = \begin{cases} 1, & 0 < t \leq 3/12 \\ 2, & 3/12 < t \leq 5/12 \\ 3, & 5/12 < t \leq 6/12 \\ 4, & 6/12 < t \leq 8/12 \\ 5, & 8/12 < t \leq 10/12 \\ 6, & 10/12 < t \leq 1 \end{cases} \tag{6}$$

$$F_Y^{-1}(t) = \begin{cases} 1, & 0 < t \leq 2/12 \\ 2, & 2/12 < t \leq 4/12 \\ 3, & 4/12 < t \leq 6/12 \\ 4, & 6/12 < t \leq 7/12 \\ 5, & 7/12 < t \leq 9/12 \\ 6, & 9/12 < t \leq 1 \end{cases} \tag{7}$$

Distribution functions

- Again, the distance between the inverse c.d.f.s, and, hence, between the corresponding random variables, can be computed as the maximal absolute deviation between them,

$$\mu(X, Y) = \sup_t |F_X^{-1}(t) - F_Y^{-1}(t)|,$$

  or as the area between their graphs.
- In fact, the area between the graphs of the c.d.f.s and the inverse c.d.f.s is one and the same, therefore formula (5) holds.

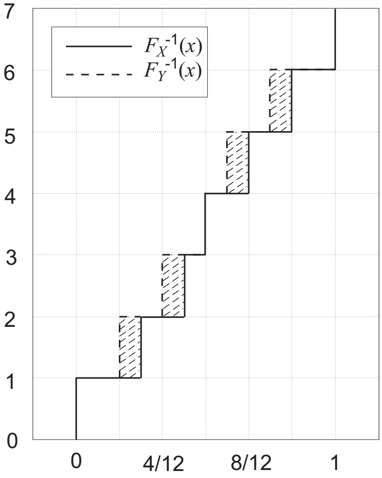

Figure: The plot shows the inverse c.d.f.s of “die X” and “die Y”. The area closed between the graphs of the two functions is shaded.

Joint distribution

- We've already considered the discrete r.v. X and Y separately, without their joint distribution. Here we will construct metrics directly using the joint distribution on the example of 2 coins.
- First, let us consider a pair of fair coins with joint probabilities as given below. The outcomes are traditionally denoted by zero and one and the joint probabilities indicate that the outcomes of the two coins are independent events.

		"coin X"	
		0	1
"coin Y"	0	1/4	1/4
	1	1/4	1/4

Table: The joint probabilities of the outcomes of two fair coins.

Joint distribution

- Both coins are fair and, therefore, they are indistinguishable if considered separately, as stand-alone random mechanisms.
- The distance between the two random variables behind the random mechanism is zero on the basis of the discussed approach. They have the same distribution functions and, consequently, all kinds of characteristics are also the same.

$\Rightarrow$ In effect, any kind of metric based on the distribution function would indicate zero distance between the two random variables.

Joint distribution

- Of course, the two random variables are not the same. They only have identical probabilistic properties.
- For instance, the conditional probability $P(X = 0|Y = 1) = 1/2$ and it follows that the events $\{X = 0, Y = 1\}$ and $\{X = 0, Y = 0\}$ may both occur if we observe realizations of the pair.
- If we would like to measure the distance between the random variables themselves, we need a different approach than the ones described above. If the random variables are defined on the same probability space (i.e. if we know their joint distribution), then we can take advantage of the additional information.

Joint distribution

- One way to calculate the distance between the two random variables is through an absolute moment of the difference $X - Y$, for example,

$$\mu(X, Y) = E|X - Y|. \tag{8}$$

- A simple calculation shows that $\mu(X, Y) = 1/2$ for the joint distribution in the table on slide 23.

Joint distribution

- The joint distribution of a pair of random variables (X, Y) provides a complete description of the probabilistic properties of the pair.
- We can compute the one-dimensional distribution functions; that is, we know the probabilistic properties of the variables if viewed on a stand-alone basis, and we also know the dependence between X and Y.
- If we keep the one-dimensional distributions fixed and change the dependence only, does the distance between the random variables change?

  The answer is affirmative and we can illustrate it with the metric (8) using the joint distribution in the table already given on slide 23.

Joint distribution

- The absolute difference $|X - Y|$ in this case may take only two values — zero and one.
- The mean $E|X - Y|$ can increase or decrease depending on the probabilities of the two outcomes.
- We have to keep in mind that the one-dimensional probabilities should remain unchanged, i.e. the sums of the numbers in the rows and the columns should be fixed to 1/2.
- Now it is easy to see how the probability mass has to be reallocated so that we obtain the minimal $E|X - Y|$ — we have to increase the probability of the outcome $(X = 0, Y = 0)$ and $(X = 1, Y = 1)$ and reduce the probabilities of the other two outcomes.

- We arrive at the conclusion that the minimal $E|X - Y|$ is attained at the joint distribution given in table below.
- The minimal $E|X - Y|$ is called the minimal metric.

		"coin X"	
		0	1
"coin Y"	0	1/2	0
	1	0	1/2

Table: The joint probabilities of the outcomes of two fair coins yielding the minimal $E|X - Y|$.

- The minimal $E|X - Y|$ in this case is equal to zero. Because the joint distribution implies that the only possible outcomes are $(X = 0, Y = 0)$ and $(X = 1, Y = 1)$ which means that the two random variables cannot be distinguished. In all states of the world with non-zero probability, they take identical values.

Joint distribution

- The exercise of finding the maximal $E|X - Y|$ is an alternative to finding the minimal metric.
- Now we have to increase the probability of $(X = 0, Y = 1)$ and $(X = 1, Y = 0)$ and reduce the probabilities of the other two outcomes.
- Finally, we find that the maximal $E|X - Y|$ is attained at the joint distribution given in the table below. The maximal $E|X - Y|$ is called the maximal distance because it does not have metric properties.

		"coin X"	
		0	1
"coin Y"	0	0	1/2
	1	1/2	0

Table: The joint probabilities of the outcomes of two fair coins yielding the maximal $E|X - Y|$.

Joint distribution

- Note that in latter case the only possible outcomes are $(X = 0, Y = 1)$ and $(X = 1, Y = 0)$ and thus the two random variables are, in a certain sense, “maximally distinct”.
- There is not a single state of the world with non-zero probability in which the two random variables take identical values.

Joint distribution

- When considering two fair coins, we checked that the minimal $E|X - Y|$ is equal to zero.
- If the one-dimensional distribution of the coins were not the same then we would not obtain a zero distance from the minimal metric.
- For example, let us consider two coins, “coin U” and “coin V” with joint probabilities as given in the table below.

		“coin U”	
		0	1
“coin V”	0	3/20	7/20
	1	2/20	8/20

Table: The joint probabilities of the outcomes “coin U” and “coin V”

$\Rightarrow$ It becomes clear that “coin V” is fair, while “coin U” is unfair — the event “0” happens with probability 5/20 and the event “1” with probability 15/20.

Joint distribution

- The same arguments as in the fair-coin example show that the minimal $E|U - V|$ and the maximal $E|U - V|$ are achieved at the joint distributions given in the tables below.

		"coin U"	
		0	1
"coin V"	0	1/4	1/4
	1	0	1/2

Table: The joint probabilities yielding minimal $E|U - V|$

		"coin U"	
		0	1
"coin V"	0	0	1/2
	1	1/4	1/4

Table: The joint probabilities yielding maximal $E|U - V|$

Joint distribution

- There is a remarkable relationship between minimal metrics and the metrics based on the distribution functions.
- For example, the metric (5) applied to the one-dimensional distributions of the two coins U and V yields exactly 1/4, which is also the value of the minimal $E|U - V|$.

Primary, simple, and compound metrics

Here we'll revisit the ideas considered in the previous section at a more advanced level with continuous random variables' examples.

Important topics will be discussed such as:

- Axiomatic construction of probability metrics
- Distinction between the three classes of primary, simple, and compound metrics
- Minimal and maximal distances

Axiomatic construction

- Generally, a metric, or a metric function, defines the distance between elements of a given set.
- Metrics are introduced axiomatically; that is, any function which satisfies a set of axioms is called a metric.
- A functional which measures the distance between random quantities is called a probability metric.
- These random quantities can be random variables, such as the daily returns of equities, the daily change of an exchange rate, etc., or stochastic processes, such as a price evolution 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 random variables only.

Axiomatic construction

- There are special properties that should be satisfied in order for the functional to be called a probability metric.
- These special properties are the axioms which constitute the building blocks behind the axiomatic construction:

1. The first axiom states that the distance between a random quantity and itself should be zero while in general, it is a non-negative number,

   Property 1. $\mu(X,Y) \geq 0$ for any X, Y and $\mu(X,X) = 0$

   Any other requirement will necessarily result in logical inconsistencies.

Axiomatic construction

2. The second axiom demands that the distance between X and Y should be the same as the distance between Y and X and is referred to as the symmetry axiom,

   Property 2. $\mu(X, Y) = \mu(Y, X)$ for any X, Y

3. The third axiom is essentially an abstract version of the triangle inequality — the distance between X and Y is not larger than the sum of the distances between X and Z and between Z and Y,

   Property 3. $\mu(X, Y) \leq \mu(X, Z) + \mu(Z, Y)$ for any X, Y, Z

$\Rightarrow$ Any functional satisfying Property 1, 2, and 3 is called probability metric.

Primary metrics

- Suppose that X and Y stand for the random returns of 2 equities.
- Then what is meant by X being the same or coincident to Y? It is that X and Y are indistinguishable in a certain sense. This sense could be to the extent of a given set of characteristics of X and Y.
- For example, X is to be considered indistinguishable to Y if their expected returns and variances are the same. Therefore, a way to define the distance between them is through the distance between the corresponding characteristics, i.e., how much their expected returns and variances deviate.
- One example is

$$\mu(X, Y) = |EX - EY| + |\sigma^2(X) - \sigma^2(Y)|$$

Such probability metrics are called primary metrics, and they imply the weakest form of sameness.

Primary metrics may be relevant in the following case:

- Suppose that we adopt the normal distribution to model the returns of two equities X and Y.
- We estimate the mean of equity X to be larger than the mean of equity Y, $EX > EY$. We may want to measure the distance between X and Y in terms of their variances only because if $|\sigma^2(X) - \sigma^2(Y)|$ turns out to be zero, then, on the basis of our assumption, we conclude that we prefer X to Y.
- Certainly this conclusion may turn out to be totally incorrect because the assumption of normality may be completely wrong.

Common examples of primary metrics include,

1. The engineer's metric

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

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

Simple metrics

- From probability theory we know that a random variable X is completely described by its cumulative distribution function.
  $\Rightarrow$ If we know the distribution function, then we can calculate all kinds of probabilities and characteristics.
- In the case of equity returns, we can compute the probability of the event that the return falls below a given target or the expected loss on condition that the loss is below a target.
- Therefore, zero distance between X and Y can imply complete coincidence of the distribution functions $F_X(x)$ and $F_Y(x)$ of X and Y and therefore, a stronger form of sameness.
- Probability metrics which essentially measure the distance between the corresponding distribution functions are called simple metrics.

Simple metrics

In the case of continuous random variables, is it possible to determine how many characteristics we need to include so that the primary metric turns essentially into a simple metric?

- In contrast to the discrete case, the question does not have a simple answer.
- Generally, a very rich set of characteristics will ensure that the distribution functions coincide. Such a set is, for example, the set of all moments $Eg(X)$ where the function g is a bounded, real-valued continuous function.
- Clearly, this is without any practical significance because this set of characteristics is not denumerable; that is, it contains more characteristics than the natural numbers.
- Nevertheless, this argument shows the connection between the classes of primary and simple metrics.

Common examples of simple metrics are stated below:

1. The Kolmogorov metric

$$\rho(X, Y) := \sup_{x \in \mathbb{R}} |F_X(x) - F_Y(x)| \tag{9}$$

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.

Figure on the next slide illustrates the Kolmogorov metric. $|F_X(x) - F_Y(x)|$, as a function of x.

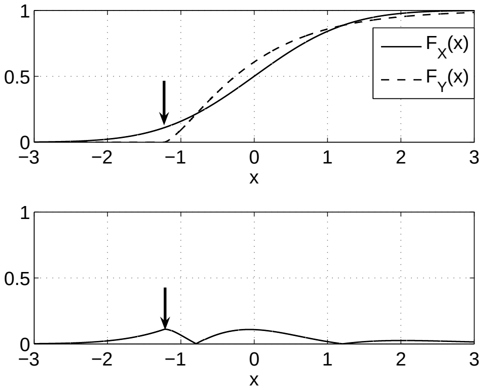

Figure: Illustration of the Kolmogorov metric. The bottom plot shows the absolute difference between the two c.d.f.s plotted on the top plot, $|F_X(x) - F_Y(x)|$, as a function of x. The arrow indicates where the largest absolute difference is attained.

Simple metrics - The Kolmogorov metric

If the r.v. X and Y describe the return distribution of 2 common stocks, then the Kolmogorov metric has the following interpretation.

- The distribution function $F_X(x)$ is 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 $\rho(X, Y)$ is the maximum deviation between the two probabilities that can be attained by varying the loss level x. If $\rho(X, Y) = 0$, 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.

2. The Lévy metric

$$\mathbf{L}(X,Y) := \inf_{\epsilon>0}\{F_X(x-\epsilon)-\epsilon \le F_Y(x) \le F_X(x+\epsilon)+\epsilon,\ \forall x \in \mathbb{R}\} \quad (10)$$

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 $\rho(X, Y) = 0$ and $\mathbf{L}(X, Y) = 0$ imply coincidence of the distribution functions $F_X(x)$ and $F_Y(x)$.
- 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

$$\mathbf{L}(X, Y) \leq \rho(X, Y) \tag{11}$$

Simple metrics - The Lévy metric

- Suppose that X is a random variable describing the return distribution of a portfolio of stocks and Y is a deterministic benchmark with a return of 2.5% ($Y = 2.5\%$).
- Assume also that the portfolio return has a normal distribution with mean equal to 2.5% and a volatility σ.
- 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,

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

- 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

$$\kappa(X, Y) := \int_{\mathbb{R}} |F_X(x) - F_Y(x)| dx. \quad (12)$$

where X and Y are random variables 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. describing the return distribution of 2 common stocks. Then $F_X(x)$ 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.

Simple metrics - The Kantorovich metric

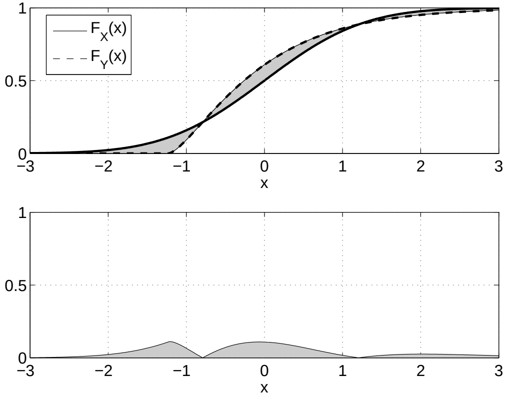

Figure: Illustration of the Kantorovich metric. The bottom plot shows the absolute difference between the two c.d.f.s plotted on the top plot. The Kantorovich metric equals the shaded area.

Simple metrics - The Kantorovich metric

- 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 $\kappa(X, Y)$ but only to a small extent.

4. The L_p-metrics between distribution functions

$$\theta_p(X,Y) := \left(\int_{-\infty}^{\infty} |F_X(x) - F_Y(x)|^p \dot{x} \right)^{1/p}, \ p \geq 1. \qquad (13)$$

where X and Y are random variables with finite mathematical expectation, $EX < \infty$ and $EY < \infty$.

- The financial interpretation of $\theta_p(X, Y)$ is similar to the interpretation of the Kantorovich metric, which appears as a special case, $\kappa(X, Y) = \theta_1(X, Y)$.
- The metric $\theta_p(X, Y)$ is an aggregate metric of the difference between the probabilities that X and Y lose more than the level x.
- The power p makes the smaller contributors to the total sum of the Kantorovich metric become even smaller contributors to the total sum in (13).
- Thus, as p increases, only the largest absolute differences $|F_X(x) - F_Y(x)|$ start to matter. At the limit, as p approaches infinity, only the largest difference $|F_X(x) - F_Y(x)|$ becomes significant and the metric $\theta_\infty(X, Y)$ turns into the Kolmogorov metric.

5. The uniform metric between inverse distribution functions

$$\mathbf{W}(X,Y) = \sup_{0<t<1} |F_X^{-1}(t) - F_Y^{-1}(t)| \tag{14}$$

where $F_X^{-1}(t)$ is the inverse of the distribution function of the random variable X.

The uniform metric between inverse distribution functions has the following financial interpretation.

- Suppose that X and Y describe the return distribution of 2 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 $(1-t)100\%$.
- 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.

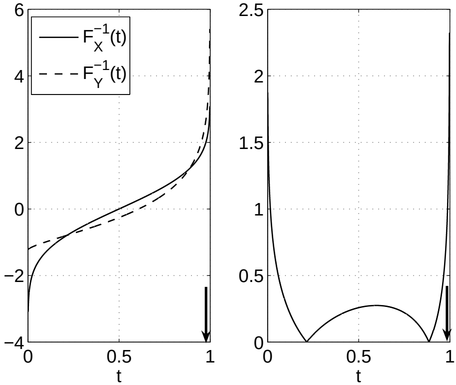

Figure: Illustration of the uniform metric between inverse distribution functions. The right plot shows the absolute difference between the two inverse c.d.f.s plotted on the left plot. The arrow indicates where the largest absolute difference is attained. Note that the inverse c.d.f.s plotted here correspond to the c.d.f.s on slide 45 (Kolmogorov metric).

6. The L_p-metrics between inverse distribution functions

$$\ell_p(X,Y) := \left(\int_0^1 |F_X^{-1}(t) - F_Y^{-1}(t)|^p \underset{.}{\mathrm{t}} \right)^{1/p}, \; p \geq 1. \qquad (15)$$

where X and Y are random variables with finite mathematical expectation, $EX < \infty$ and $EY < \infty$ and $F_X^{-1}(t)$ is the inverse of the distribution function of the random variable X.

- The metric $\ell_1(X, Y)$ is also known as first difference pseudomoment as well as the average metric in the space of distribution functions because $\ell_1(X, Y) = \theta_1(X, Y)$.
- Another notation used for this metric is $\kappa(X, Y)$, note that $\theta_1(X, Y) = \kappa(X, Y)$. This special case is called the Kantorovich metric because great contributions to the properties of $\ell_1(X, Y)$ were made by Kantorovich in 1940s.

We provide another interpretation of the Kantorovich metric arising from equation (15).

- Suppose that X and Y are r.v. describing the return distribution of 2 common stocks. We explained that the VaRs of X and Y at confidence level $(1-t)100\%$ are equal to $-F_X^{-1}(t)$ and $-F_Y^{-1}(t)$ respectively.
- Therefore, the metric

$$\ell_1(X,Y) = \int_0^1 |F_X^{-1}(t) - F_Y^{-1}(t)|dt$$

equals the sum of the absolute differences between the VaRs of X and Y across all confidence levels.
- In effect, it provides aggregate information about the deviations between the VaRs of X and Y for all confidence levels.

Simple metrics

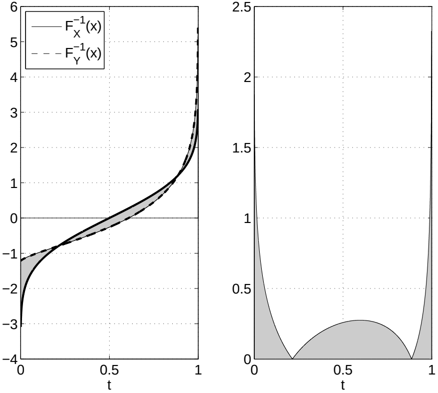

Figure: Illustration of the $\ell_1(X, Y)$ metric. The right plot shows the absolute difference between the two inverse c.d.f.s plotted on the left plot. The $\ell_1(X, Y)$ metric equals to the the largest absolute difference between the two densities, shown as shaded area.

The power p in equation (15) acts in the same way as in the case of $\theta_p(X, Y)$:

- The smaller contributors to the sum in $\ell_1(X, Y)$ become even smaller contributors to the sum in $\ell_p(X, Y)$.
- Thus, as p increases, only the larger absolute differences between the VaRs of X and Y across all confidence levels become significant in the total sum. The larger differences are in the tails of the two distributions.

  Therefore, the metric $\ell_p(X, Y)$ accentuates on the deviations between X and Y in the zone of the extreme profits or losses.

$\Rightarrow$ At the limit, as p approaches infinity, only the largest absolute differences matter and the $\ell_p(X, Y)$ metric turns into the uniform metric between inverse c.d.f.s $\mathbf{W}(X, Y)$.

7. The uniform metric between densities

$$\ell(X, Y) := \sup_{x \in \mathbb{R}} |f_X(x) - f_Y(x)| \qquad (16)$$

where $f_X(x) = F_X'(x)$ is the density of the random variable X.

Simple metrics

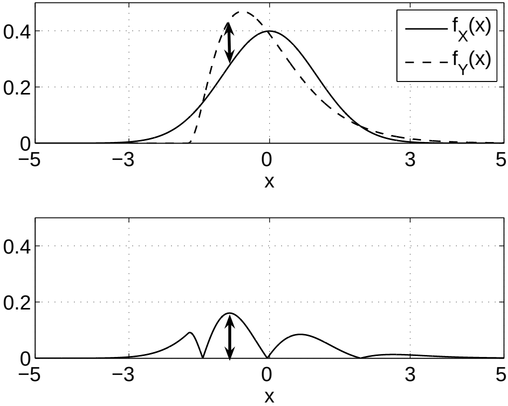

Figure: Illustration of the uniform metric between densities. The bottom plot shows the absolute difference between the two densities plotted on the top plot. The arrow indicates where the largest absolute difference is attained.

- The uniform metric between densities can be interpreted through the link between the density function and the c.d.f.
- The probability that X belongs to a small interval $[x, x + \Delta_x]$, where $\Delta_x > 0$ is small number, can be represented approximately as

$$P(X \in [x, x + \Delta_x]) \approx f_X(x).\Delta_x.$$

- Suppose that X and Y describe the return distribution of 2 common stocks.
- Then the difference between the densities $f_X(x) - f_Y(x)$ can be viewed as a quantity approximately proportional to the difference between the probabilities that X and Y realize a return belonging to the small interval $[x, x + \Delta_x]$,

$$P(X \in [x, x + \Delta_x]) - P(Y \in [x, x + \Delta_x]).$$

- Thus, the largest absolute difference between the two density functions is attained at such a return level x that the difference between the probabilities of X and Y gaining return $[x, x + \Delta_x]$ is largest in absolute value.
- Just as in the case of the Kolmogorov metric, the value of x for which the maximal absolute difference between the densities is attained is close to the mean return. Therefore, the metric $\ell(X, Y)$ is not sensitive to extreme losses or profits.

- The total variation metric

$$\sigma(X,Y) = \sup_{\text{all events } A} |P(X \in A) - P(Y \in A)| \tag{17}$$

If the random variables X and Y have densities $f_X(x)$ and $f_Y(x)$, then the total variation metric can be represented through the area closed between the graphs of the densities,

$$\sigma(X,Y) = \frac{1}{2}\int_{-\infty}^{\infty} |f_X(x) - f_Y(x)| dx. \tag{18}$$

- Suppose that X and Y describe the return distribution of 2 common stocks. We can calculate the probabilities $P(X \in A)$ and $P(Y \in A)$ where A is an arbitrary event.
- For example, A can be the event that the loss exceeds a given target x, or that the loss is in a given bound (x%, y%), or in an arbitrary unions of such bounds. The total variation metric is the maximum absolute difference between these probabilities.
- The reasoning is very similar to the one behind the interpretation of the Kolmogorov metric.

- The principal difference from the Kolmogorov metric is that in the total variation metric, we do not fix the events to be only of the type “losses exceed a given target x”.
- Instead, we calculate the maximal difference by looking at all possible types of events. Therefore, the general relationship between the two metrics is

$$\rho(X, Y) \leq \sigma(X, Y). \tag{19}$$

$\Rightarrow$ If any of these metrics turn into zero, then it follows that the distribution functions of the corresponding random variables coincide.

Compound metrics

- The coincidence of distribution functions is stronger than coincidence of certain characteristics. But there is a stronger form of identity than coincidence of distribution functions, which is actually the strongest possible.
- Consider the case in which no matter what happens, the returns of equity 1 and equity 2 are identical. As a consequence, their distribution functions are the same because the probabilities of all events of the return of equity 1 are exactly equal to the corresponding events of the return of equity 2.
- This identity is also known as almost everywhere identity because it considers all states of the world which happen with non-zero probability.
- The probability metrics which imply the almost everywhere identity are called compound metrics.

Common examples of compound metrics are stated below:

1). The p-average compound metric

$$\mathcal{L}_p(X,Y) = (E|X-Y|^p)^{1/p}, \; p \geq 1 \tag{20}$$

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

Compound metrics - The p-average compound metric

- From a financial viewpoint, 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. If p is equal to two, we obtain

$$\mathcal{L}_2(X, Y) = \sqrt{E(X - Y)^2}$$

which is a quantity very similar to the tracking error between the two portfolios.

2). The Ky Fan metric

$$K(X,Y) := \inf\{\epsilon > 0 : P(|X - Y| > \epsilon) < \epsilon\} \quad (21)$$

where X and Y are real-valued random variables.

The Ky Fan metric has an important application in theory of probability as it metrizes convergence in probability of real-valued random variables.

Compound metrics - The Ky Fan metric

- 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

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

concerns the event that either the portfolio will outperform the benchmark by ϵ or it will underperform the benchmark by ϵ.
- Therefore, the quantity 2ϵ can be interpreted as the width of a performance band.
- The probability $1 - P(|X - Y| > \epsilon)$ is actually the probability that the portfolio stays within the performance band, i.e. it does not deviate from the benchmark more than ϵ in an upward or downward direction.

Compound metrics - The Ky Fan metric

- As the width of the performance band decreases, the probability $P(|X - Y| > \epsilon)$ increases because the portfolio returns will be more often outside a smaller band.
- 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

$$\Theta_p(X,Y) = \left(\int_{-\infty}^{\infty} \tau^p(t;X,Y)dt \right)^{1/p}, \ p \geq 1 \qquad (22)$$

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

The function $\tau(t; X, Y)$ can be interpreted in the following way.

- Suppose that X and Y describe the return distributions of 2 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.

$\Rightarrow$ Therefore, 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 $p = 1$, the Birnbaum-Orlicz compound metric sums all probabilities of this type for all values of the performance divide t.
- Thus, it is an aggregate measure of the deviations in the relative performance of X and Y. In fact, it is exactly equal to the mean absolute deviation,

$$\Theta_1(X, Y) = E|X - Y| = \mathcal{L}_1(X, Y).$$

Minimal and maximal metrics

From the discussion above, three classes of probability metrics are interrelated: they are contained in one another.

Primary metrics can be “enriched” so that they turn into simple metrics by the following process.

- Suppose that we have a list of characteristics which defines the primary metric.
- Then we start adding additional characteristics which cannot be expressed in any way by means of the ones currently in the list. Assume that this process continues indefinitely, until we exhaust all possible characteristics.
- The primary metric obtained by means of the set of all possible characteristics is actually a simple metric, as we end up with coincident distribution functions.

Minimal and maximal metrics

- For instance, assume that we have a compound metric. It is influenced not only by the distribution functions but also by the dependence between the random variables.
- Is it possible to construct a simple metric on the basis of it? The answer is positive and the simple metric is built by constructing the minimal metric:

  ❏ Choose two random variables X and Y.
  ❏ Compute the distances between all possible random variables having the same distribution as the ones selected using the compound metric.
  ❏ Set the minimum of these distances to be the distance between the random variables X and Y.

  The result is a simple metric because due to the minimization, we remove the influence on the dependence structure and only the distribution functions remain. By this process, we associate a simple metric to any compound metric.

- The minimal metrics have an important place in the theory of probability metrics and there is notation reserved for them.
- Denote by μ the selected compound metric. The functional $\hat{\mu}$ defined by the equality

$$\hat{\mu}(X, Y) := \inf\{\mu(\widetilde{X}, \widetilde{Y}) : \widetilde{X} \stackrel{d}{=} X,\ \widetilde{Y} \stackrel{d}{=} Y\} \qquad (23)$$

is said to be the minimal metric with respect to μ.

Minimal and maximal metrics

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

- For example, the L_p metrics between distribution functions and inverse distribution functions defined in (13) and (15) are minimal metrics with respect to the p-average compound metric (20) and the Birnbaum-Orlicz compound metric (22),

$$\ell_p(X,Y) = \hat{\mathcal{L}}_p(X,Y)$$
$$\theta_p(X,Y) = \hat{\Theta}_p(X,Y).$$

- The Kolmogorov metric (9) can be represented as a special case of the simple metric θ_p, $\rho(X,Y) = \theta_\infty(X,Y)$ and, therefore, it also arises as a minimal metric

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

Minimal and maximal metrics

But 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, $\Theta_1(X, Y)$ is protominimal to the Kantorovich metric $\kappa(X, Y)$.

  $\Rightarrow$ Not all simple metrics have protominimal ones and, also, some simple metrics have several protominimal ones.

- The definition of the minimal metric (23) shows that the compound metric and the minimal metric relative to it are related by the inequality

$$\hat{\mu}(X, Y) \leq \mu(X, Y).$$

Minimal and maximal metrics

We can find an upper bound to the compound metric by a process very similar to finding the minimal metric.

- We choose two random variables X and Y and compute the distances by means of the compound metric between all possible random variables having the same distribution as the ones selected.
- Then we set the maximum of these distances to be the needed upper bound. Naturally, this upper bound is called maximal metric. It is denoted by

$$\check{\mu}(X, Y) := \sup\{\mu(\widetilde{X}, \widetilde{Y}) : \widetilde{X} \stackrel{d}{=} X,\ \widetilde{Y} \stackrel{d}{=} Y\} \qquad (24)$$

- Thus, we can associate a lower and an upper bound to each compound metric,

$$\hat{\mu}(X, Y) \leq \mu(X, Y) \leq \check{\mu}(X, Y).$$

- It turns out that the maximal distance is not a probability metric because the identity property may not hold, $\breve{\mu}(X, X) > 0$, as it is an upper bound to the compound metric $\mu(X, Y)$.
- Functionals which satisfy only Property 2 and Property 3 from the defining axioms of probability metrics are called moment functions. Therefore, the maximal metric is a moment function.

Minimal and maximal metrics

We illustrate the notions of minimal and maximal metrics further:

- Suppose that the pair of random variables (X, Y) has some bivariate distribution with zero-mean normal marginals, $X \in N(0, \sigma_X^2)$, $Y \in N(0, \sigma_Y^2)$. The particular form of the bivariate distribution is insignificant.
- Let us calculate the minimal and the maximal metrics of the 2-average compound metric $\mathcal{L}_2(X, Y) = (E(X - Y)^2)^{1/2}$.
- In fact, the compound metric $\mathcal{L}_2(X, Y)$ stands for the standard deviation of the difference $X - Y$. The variance of the difference, σ_{X-Y}^2, can be calculated explicitly,

$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2 - 2\sigma_X \sigma_Y \operatorname{corr}(X, Y)$$

  where $\operatorname{corr}(X, Y)$ denotes the correlation coefficient between X and Y.

Minimal and maximal metrics

- Holding the one-dimensional distributions fixed and varying the dependence model, or the copula function, in this case means that we hold fixed the variances σ_X^2 and σ_Y^2 and we vary the correlation $\mathrm{corr}(X, Y)$.
- This is true because the one-dimensional normal distributions are identified only by their variances. Recall that the absolute value of the correlation coefficient is bounded by one,

$$-1 \leq \mathrm{corr}(X, Y) \leq 1,$$

and, as a result, the lower and upper bounds of the variance σ_{X-Y}^2 are

$$\sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y \leq \sigma_{X-Y}^2 \leq \sigma_X^2 + \sigma_Y^2 + 2\sigma_X\sigma_Y.$$

Minimal and maximal metrics

- Note that the bounds for the correlation coefficient are not tied to any sort of distributional hypothesis and are a consequence of a very fundamental inequality in mathematics known as the Cauchy-Bunyakovski-Schwarz inequality.
- As a result, we obtain bounds for the standard deviation of the difference $X - Y$ which is the 2-average compound metric,

$$|\sigma_X - \sigma_Y| \leq \mathcal{L}_2(X, Y) \leq \sigma_X + \sigma_Y.$$

- We have followed strictly the process of obtaining minimal and maximal metrics. Therefore, we conclude that, in the setting of the example,

$$\hat{\mathcal{L}}_2(X, Y) = |\sigma_X - \sigma_Y| \qquad \text{and} \qquad \check{\mathcal{L}}_2(X, Y) = \sigma_X + \sigma_Y.$$

- An example of an explicit expression for a maximal metric is the p-average maximal distance

$$\check{\mathcal{L}}_p(X, Y) = \left(\int_0^1 (F_X^{-1}(t) - F_Y^{-1}(1-t))^p dt \right)^{1/p}, \; p \geq 1 \qquad (25)$$

where $F_X^{-1}(t)$ is the inverse of the distribution function of the random variable X.

Svetlozar T. Rachev, Stoyan Stoyanov, and Frank J. Fabozzi
Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization: The Ideal Risk, Uncertainty, and Performance Measures
John Wiley, Finance, 2007.

Chapter 3.