Technical Appendix

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

Technical Appendix

- The distance between various objects, such as vectors, matrices, functions, etc., are measured by means of special functions called metrics.
- The notion of a metric function, usually denoted by $\rho(x, y)$, is actually very fundamental. It defines the distance between elements of a given set.
- The most common example is the Euclidean metric,

$$\rho(x, y) = \sqrt{\sum_{i=1}^{n} (x_i^2 - y_i^2)}$$

where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ are vectors in $\mathbb{R}^n$, which has a very intuitive meaning in the real plane. It calculates the length of the straight line connecting the two points x and y.

Metric functions are defined through a number of axioms.

- A set S is said to be a metric space endowed with the metric ρ if ρ is a mapping from the product $S \times S$ to $[0, \infty)$ having the following properties for each $x,\ y,\ z \in S$

$$\begin{array}{ll} \textit{Identity property}: & \rho(x, y) = 0 \iff x = y \\ \textit{Symmetry}: & \rho(x, y) = \rho(y, x) \\ \textit{Triangle inequality}: & \rho(x, y) \leq \rho(x, z) + \rho(z, y) \end{array}$$

Technical Appendix

- An example of a metric space is the n-dimensional vector space $\mathbb{R}^n$ with the metric

$$\rho(x,y) = ||x-y||_p = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p}, \; p \geq 1.$$

  Clearly, the Euclidean metric appears when $p = 2$.
- The same ideas behind the definition of a metric function ρ are used in the definition of probability metrics.

- The first axiom, called the identity property, is a reasonable requirement. In the theory of probability metrics, we distinguish between two varieties,

$$\begin{array}{ll} \text{ID.} & \mu(X,Y) \geq 0 \text{ and } \mu(X,Y) = 0, \text{ if and only if } X \sim Y \\ \widetilde{\text{ID}}. & \mu(X,Y) \geq 0 \text{ and } \mu(X,Y) = 0, \text{ if } X \sim Y \end{array}$$

- The notation $X \sim Y$ denotes that X is equivalent to Y. The meaning of equivalence depends on the type of metrics.
- If we consider compound metrics, then the equivalence is in almost sure sense. If we consider simple metrics, then $\sim$ means equality of distribution and, finally, if we consider primary metrics, then $\sim$ stands for equality of some characteristics of X and Y. The axiom $\widetilde{\text{ID}}$ is weaker than ID.

- The symmetry axiom makes sense in the general context of calculating distances between elements of a space,

$$\text{SYM.} \quad \mu(X, Y) = \mu(Y, X)$$

- The third axiom is the triangle inequality,

$$\text{TI.} \quad \mu(X, Y) \leq \mu(X, Z) + \mu(Z, Y) \text{ for any } X, Y, Z$$

- The triangle inequality is important because it guarantees, together with ID, that μ is continuous in any of the two arguments. This nice mathematical property appears as a result of the consequence of TI,

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

- Observe that if the distance between Z and Y as measured by $\mu(Z, Y)$ is small, so is the left hand-side of the inequality above. That is, intuitively, small deviations in the second argument of the functional $\mu(X, \cdot)$ correspond to small deviations in the functional values. The same conclusion holds for the first argument.
- The triangle inequality can be relaxed to the more general form called triangle inequality with parameter K,

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

Notice that the traditional version TI appears when $K = 1$.

- Notice that in the two versions of the triangle inequality, the statement that the inequality holds *for any* X, Y, Z is not very precise.
- In fact, we are evaluating the functional μ for a pair of random variables, for example (X, Y), and μ shows the distance between the random variables in the pair.
- The pair cannot be dismantled to its constituents because the random variables X and Y are coupled together by their dependence structure and if μ is a compound functional, then how X and Y are coupled is important.
- Therefore, the triangle inequality holds for the three pairs (X, Y), (X, Z), and (Y, Z).

- As matter of fact, the three pairs cannot be arbitrary. Suppose that we choose the first pair (X, Y) and the second pair (X, Z); that is, we fix the dependence between X and Y in the first pair, and X and Z in the second pair. Under these circumstances, it is obvious that the dependence between Z and Y cannot be arbitrary but should be consistent with the dependence of the chosen pairs (X, Y) and (X, Z).
- *But then, is there any freedom in the choice of the pair (Z, Y)? Do these arguments mean that by choosing the two pairs (X, Y) and (X, Z) we have already fixed the pair (Z, Y)?*
- It turns out that the pair (Z, Y) is not fixed by the choice of the other two ones. We are free to choose the dependence in the pair (Z, Y) as long as we do not break the following consistency rule.

Consistency rule: The three pairs of random variables (X, Y), (X, Z), and (Z, Y) 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.

Here is an example illustrating the consistency rule.

- Let us choose a metric μ. Suppose that we would like to verify if the triangle inequality holds by choosing three pairs of random variables.
- The distribution of all pairs is assumed to be bivariate normal with zero mean, $(X, Y) \in N(0, \Sigma_1)$, $(X, Z) \in N(0, \Sigma_2)$, and $(Z, Y) \in N(0, \Sigma_3)$ where the covariance matrices are given by

$$\Sigma_1 = \begin{pmatrix} 1 & 0.99 \\ 0.99 & 1 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 1 & 0.99 \\ 0.99 & 1 \end{pmatrix}, \quad \Sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

- Do these three pairs satisfy the consistency rule? Note that the correlation between X and Y is very strongly positive, $\mathrm{corr}(X, Y) = 0.99$. The correlation between X and Z is also very strongly positive, $\mathrm{corr}(X, Z) = 0.99$.

- *Then, is it possible that Z and Y be independent?*
  The answer is no because, under our assumption, when X takes a large positive value, both X and Y take large positive values which implies strong dependence between them.
- The consistency rule states that the dependence between Y and Z should be such that the three pairs can be consistently embedded in a three-dimensional vector.
  Then, can we find a value for the correlation between Z and Y so that this becomes possible?

- We can find a partial answer to the last question by searching for a consistent three-dimensional normal distribution such that its two dimensional projections are the given bivariate normal distributions.
- That is, we are free to choose the correlation between Z and Y, $\mathrm{corr}(Z, Y) = \sigma_{ZY}$, on condition that the matrix

$$\begin{pmatrix} 1 & 0.99 & 0.99 \\ 0.99 & 1 & \sigma_{ZY} \\ 0.99 & \sigma_{ZY} & 1 \end{pmatrix}$$

is a valid covariance matrix, i.e. it should be positive definite.
- For this particular example, it can be calculated that the consistency condition holds if $\sigma_{ZY} \geq 0.9602$.

Combinations of the defining axioms considered above imply different properties and, consequently, the functionals defined by them have specific names. If a functional μ satisfies,

- ID, SYM and TI, then μ is called probability metric
- $\widetilde{\mathrm{ID}}$, SYM, TI, then μ is called called probability semimetric
- ID, SYM, $\widetilde{\mathrm{TI}}$, then μ is called called probability distance
- $\widetilde{\mathrm{ID}}$, SYM, $\widetilde{\mathrm{TI}}$, then μ is called called probability semidistance

In financial applications in particular, the symmetry axiom is not important and is better to omit it.

Thus, we extend the treatment of these axioms 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. That is, if μ satisfies,

- ID and TI, then μ is called probability quasi-metric
- $\widetilde{\mathrm{ID}}$, TI, then μ is called called probability quasi-semimetric
- ID, $\widetilde{\mathrm{TI}}$, then μ is called called probability quasi-distance
- $\widetilde{\mathrm{ID}}$, $\widetilde{\mathrm{TI}}$, then μ is called called probability quasi-semidistance

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

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

$$K_H := \sup_{t>0} \frac{H(2t)}{H(t)} < \infty \tag{1}$$

  which is known as Orlicz's condition.
- There is a general result which states that if ρ 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 (1) by $\mathcal{H}$.

- The engineer's distance

$$\mathbf{EN}(X, Y; H) := H\left(|EX - EY|\right), \quad H \in \mathcal{H} \tag{2}$$

where the random variables X and Y have finite mathematical expectation, $E|X| < \infty$, $E|Y| < \infty$.

1. The Kantorovich distance

$$\ell_H(X,Y) := \int_0^1 H(|F_X^{-1}(t) - F_Y^{-1}(t)|)dt, \quad H \in \mathcal{H} \qquad (3)$$

where the random variables 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 in (15) in the Lecture. Note that L_p metric between inverse distribution functions, $\ell_p(X,Y)$, can be slightly extended to

$$\ell_p(X,Y) := \left(\int_0^1 |F_X^{-1}(t) - F_Y^{-1}(t)|^p dt\right)^{1/\min(1,1/p)}, \ p > 0. \qquad (4)$$

- Under this slight extension, the limit case $p \to 0$ appears to be the total variation metric defined in (17),

$$\ell_0(X,Y) = \sigma(X,Y) = \sup_{\text{all events } A} |P(X \in A) - P(Y \in A)|. \tag{5}$$

- The other limit case provides a relation to the uniform metric between inverse distribution functions $\mathbf{W}(X,Y)$ given by (14),

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

2. The Birnbaum-Orlicz average distance

$$\theta_H(X,Y) := \int_{\mathbb{R}} H(|F_X(x) - F_Y(x)|)dx, \quad H \in \mathcal{H} \tag{6}$$

where the random variables X and Y have finite mathematical expectation, $E|X| < \infty$, $E|Y| < \infty$.

- If we choose $H(t) = t^p,\ p \geq 1$, then $(\theta_H(X,Y))^{1/p}$ turns into the L_p metric between distribution functions, $\theta_p(X,Y)$, defined in (13). Note that L_p metric between distribution functions, $\theta_p(X,Y)$, can be slightly extended to

$$\theta_p(X,Y) := \left(\int_{-\infty}^{\infty} |F_X(x) - F_Y(x)|^p dx\right)^{1/\min(1,1/p)}, \; p > 0. \tag{7}$$

- At limit as $p \to 0$,

$$\theta_0(X,Y) := \int_{-\infty}^{\infty} I\{x : F_X(x) \neq F_Y(x)\} dx \qquad (8)$$

  where the notation $I\{A\}$ stands for the indicator of the set A. That is, the simple metric $\theta_0(X,Y)$ calculates the Lebesgue measure of the set $\{x : F_X(x) \neq F_Y(x)\}$.
- If $p \to \infty$, then we obtain the Kolmogorov metric defined in (9) in the Lecture, $\theta_\infty(X,Y) = \rho(X,Y)$.

3. The Birnbaum-Orlicz uniform distance

$$\begin{aligned}\rho_H(X,Y) &:= H(\rho(X,Y)) \\ &= \sup_{x\in\mathbb{R}} H(|F_X(x) - F_Y(x)|), \quad H \in \mathcal{H}\end{aligned} \tag{9}$$

The Birnbaum-Orlicz uniform distance is a generalization of the Kolmogorov metric.

4. The parametrized Lévy metric

$$\mathbf{L}_\lambda(X,Y) := \inf\{\epsilon > 0 : F_X(x-\lambda\epsilon) - \epsilon \leq F_Y(x) \leq F_X(x+\lambda\epsilon) + \epsilon,\ \forall x \in \mathbb{R}\} \tag{10}$$

This is a parametric extension of the Lévy metric, $\mathbf{L}(X,Y)$, defined by (10). The obvious relationship with the Lévy metric is $\mathbf{L}_1(X,Y) = \mathbf{L}(X,Y)$.

- It is possible to show that the parametric extension $\mathbf{L}_\lambda(X,Y)$ is related to the celebrated Kolmogorov metric, $\rho(X,Y)$, defined by (9) and the uniform metric between inverse distribution functions, $\mathbf{W}(X,Y)$, given by equation (14),

$$\lim_{\lambda\to 0} \mathbf{L}_\lambda(X,Y) = \rho(X,Y) \quad \text{and} \quad \lim_{\lambda\to\infty} \lambda\mathbf{L}_\lambda(X,Y) = \mathbf{W}(X,Y).$$

1. The H-average compound distance

$$\mathcal{L}_H(X,Y) := E(H(|X-Y|)), \quad H \in \mathcal{H} \tag{11}$$

- If we choose $H(t) = t^p,\ p \geq 1$, then $(\mathcal{L}_H(X,Y))^{1/p}$ turns into the p-average metric, $\mathcal{L}_p(X,Y)$, defined in (20). Note that the p-average metric can be slightly extended to

$$\mathcal{L}_p(X,Y) := (E|X-Y|^p)^{1/\min(1,1/p)}, \quad p > 0 \tag{12}$$

- At the limit, as $p \to 0$, we define

$$\mathcal{L}_0(X,Y) := P(\{w : X(w) \neq Y(w)\}) \tag{13}$$

If $p \to \infty$, then we define

$$\mathcal{L}_\infty(X,Y) := \inf\{\epsilon > 0 : P(|X-Y| > \epsilon) = 0\} \tag{14}$$

2. The Ky-Fan distance

$$\mathbf{KF}_H(X,Y) := \inf\{\epsilon > 0 : P(H(|X - Y|) > \epsilon) < \epsilon\}, \quad H \in \mathcal{H} \quad (15)$$

- A particular case of the Ky-Fan distance is the *parametric family of Ky-Fan metrics*

$$\mathbf{K}_\lambda(X,Y) := \inf\{\epsilon > 0 : P(|X - Y| > \lambda\epsilon) < \epsilon\}, \quad \lambda > 0 \quad (16)$$

- The parametric family $\mathbf{K}_\lambda(X,Y)$ has application in the theory of probability since, for each $\lambda > 0$, $\mathbf{K}_\lambda(X,Y)$ metrizes the convergence in probability. That is, if $X_1,\ldots,X_n,\ldots$ is a sequence of random variables, then

$$\mathbf{K}_\lambda(X_n,Y) \to 0 \quad \Longleftrightarrow \quad P(|X_n - Y| > \epsilon) \to 0, \text{ for any } \epsilon > 0.$$

- The parametric family $\mathbf{K}_\lambda(X, Y)$ is related to the p-average compound metric. The following relations hold,

$$\lim_{\lambda \to 0} \mathbf{K}_\lambda(X, Y) = \mathcal{L}_0(X, Y) \quad \text{and} \quad \lim_{\lambda \to \infty} \lambda \mathbf{K}_\lambda(X, Y) = \mathcal{L}_\infty(X, Y)$$

- Even though the Ky-Fan metrics imply convergence in probability, these two limit cases induce stronger convergence. That is, if $X_1, \ldots, X_n, \ldots$ is a sequence of random variables, then

$$\mathcal{L}_0(X_n, Y) \to 0 \quad \overset{\Rightarrow}{\not\Leftarrow} \quad X_n \to Y \text{ “in probability”}$$

and

$$\mathcal{L}_\infty(X_n, Y) \to 0 \quad \overset{\Rightarrow}{\not\Leftarrow} \quad X_n \to Y \text{ “in probability”}$$

3. The Birnbaum-Orlicz compound average distance

$$\boldsymbol{\Theta}_H(X, Y) := \int_{-\infty}^{\infty} H(\tau(t; X, Y))dt, \quad H \in \mathcal{H} \tag{17}$$

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

- If we choose $H(t) = t^p,\ p \geq 1$, then $(\boldsymbol{\Theta}_H(X, Y))^{1/p}$ turns into the Birnbaum-Orlicz average metric, $\boldsymbol{\Theta}_p(X, Y)$, defined in (22) in the Lecture. Note that the Birnbaum-Orlicz average metric can be slightly extended to

$$\boldsymbol{\Theta}_p(X, Y) := \left(\int_{-\infty}^{\infty} (\tau(t; X, Y))^p dt\right)^{1/\min(1,1/p)}, \quad p > 0 \tag{18}$$

- At the limit, as $p \to 0$, we define

$$\mathbf{\Theta}_0(X, Y) := \int_{-\infty}^{\infty} I\{t : \tau(t; X, Y) \neq 0\} dt \tag{19}$$

where $I\{A\}$ is the indicator of the set A. If $p \to \infty$, then we define

$$\mathbf{\Theta}_\infty(X, Y) := \sup_{t \in \mathbb{R}} \tau(t; X, Y) \tag{20}$$

4. The Birnbaum-Orlicz compound uniform distance

$$\mathbf{R}_H(X,Y) := H(\mathbf{\Theta}_\infty(X,Y)) = \sup_{t\in\mathbb{R}} H(\tau(t;X,Y)), \quad H \in \mathcal{H} \tag{21}$$

This is the compound uniform distance of the Birnbaum-Orlicz family of compound metrics.

Minimal and maximal distances

- We noted that two functionals can be associated to any compound metric $\mu(X, Y)$ — the minimal metric $\hat{\mu}(X, Y)$ and the maximal metric $\check{\mu}(X, Y)$ defined with equations (23) and (24) in the Lecture, respectively. The relationship between the three functionals is

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

- Exactly the same process can be followed in order to construct minimal and maximal distances, minimal and maximal semi-distances, minimal and maximal quasi-distances, etc. It turns out that the minimal functional

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

  is metric, distance, semi-distance or quasi-semidistance whenever $\mu(X, Y)$ is metric, distance, semi-distance or quasi-semidistance.
- The minimization preserves the essential triangle inequality with parameter $K_{\hat{\mu}} = K_{\mu}$ and also the identity property assumed for μ.

- In contrast, the maximal functional

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

does not preserve all properties of $\mu(X,Y)$ and, therefore, it is *not* a probability distance.

- In fact, the maximization does not preserve the important identity property, while the triangle inequality holds with parameter $K_{\breve{\mu}} = K_{\mu}$.
- As we noted in the lecture, functionals which satisfy properties SYM and $\widetilde{\mathrm{TI}}$ and fail to satisfy the identity property are called *moment functions*. Thus, the maximal distance is a moment function.

Minimal and maximal distances

- Many simple probability distances arise as minimal semi-distances with respect to some compound semi-distance.
- If $H \in \mathcal{H}$ is a convex function, then

$$\ell_H(X, Y) = \hat{\mathcal{L}}_H(X, Y)$$
$$\boldsymbol{\theta}_H(X, Y) = \hat{\mathbf{\Theta}}_H(X, Y)$$
$$\boldsymbol{\rho}_H(X, Y) = \hat{\mathbf{R}}_H(X, Y).$$

Minimal and maximal distances

- A very general result, which is used to obtain explicit expressions for minimal and maximal functionals such as $\hat{\mu}(X, Y)$ and $\check{\mu}(X, Y)$, is the Cambanis-Simons-Stout theorem.
- This theorem provides explicit forms of the minimal and maximal functionals 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*.
- The index ϕ is a reminder that the functional has the particular form with the ϕ function.

Minimal and maximal distances

- Then for the minimal and the maximal functionals $\hat{\mu}_\phi(X, Y)$ and $\check{\mu}_\phi(X, Y)$ we have the explicit representations,

$$\hat{\mu}_\phi(X, Y) = \int_0^1 \phi(F_X^{-1}(t), F_Y^{-1}(t))dt \qquad (22)$$

and

$$\check{\mu}_\phi(X, Y) = \int_0^1 \phi(F_X^{-1}(t), F_Y^{-1}(1-t))dt \qquad (23)$$

Minimal and maximal distances

- The function $\phi(x, y)$ is called quasi-antitone if it satisfies the following property

$$\phi(x, y) + \phi(x', y') \leq \phi(x', y) + \phi(x, y') \quad (24)$$

for any $x' > x$ and $y' > y$.

- This property is related to how the function increases when its arguments increase.
- Also, the function ϕ should satisfy the technical condition that $\phi(x, x) = 0$. There is another technical condition which is related to the random variables X and Y. The following moments should be finite, $E\phi(X, a) < \infty$ and $E\phi(Y, a) < \infty$, $a \in \mathbb{R}$.

Minimal and maximal distances

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.

How do we apply the Cambanis-Simons-Stout theorem?
There are three steps.

Step 1. Identify the function $\phi(x, y)$ from the particular form of the compound metric.

Step 2. Verify if the function $\phi(x, y)$ is quasi-antitone and whether $\phi(x, x) = 0$. This can be done by verifying first if $\phi(x, y)$ belongs to any of the examples of quasi-antitone functions given above.

Step 3. Keep in mind that whenever we have to apply the result in the theorem for particular random variables (X, Y), then the following moments should satisfy the conditions $E\phi(X, a) < \infty$ and $E\phi(Y, a) < \infty$, $a \in \mathbb{R}$. Otherwise, the corresponding metrics may explode.

Minimal and maximal distances

- Let us see how the Cambanis-Simons-Stout result is applied to the H-average compound distance $\mathcal{L}_H(X,Y)$ defined in (11).
- The compound functional has the general form,

$$\mathcal{L}_H(X,Y) = E(H(|X-Y|)), \quad H \in \mathcal{H}$$

and the function $\phi(x,y) = H(|x-y|)$, $x,y \in \mathbb{R}$. Due to the properties of the function H, $\phi(x,x) = H(0) = 0$ and, if we assume additionally that H is a convex function, we obtain that $\phi(x,y)$ is quasi-antitone.

- Applying the theorem yields the following explicit forms of the minimal and the maximal distance,

$$\hat{\mathcal{L}}_H(X,Y) = \int_0^1 H(|F_X^{-1}(t) - F_Y^{-1}(t)|)dt, \quad H \in \mathcal{H}$$

and

$$\check{\mathcal{L}}_H(X,Y) = \int_0^1 H(|F_X^{-1}(t) - F_Y^{-1}(1-t)|)dt, \quad H \in \mathcal{H}. \qquad (25)$$

- We tacitly assume that the technical conditions $E(H(|X-a|)) < \infty$ and $E(H(|Y-a|)) < \infty$, $a \in \mathbb{R}$ hold.

- There is another method of obtaining explicit forms of minimal and maximal functionals - the direct application of the celebrated Fréchet-Hoeffding inequality between distribution functions,

$$\max(F_X(x) + F_Y(y) - 1, 0) \leq P(X \leq x, Y \leq y) \leq \min(F_X(x), F_Y(y)). \tag{26}$$

- This inequality is applied to the problem of finding the minimal distance of the Birnbaum-Orlicz distance defined in (17):

$$\begin{aligned}
\boldsymbol{\Theta}_H(X,Y) &= \int_{-\infty}^{\infty} H(P(X \leq t < Y) + P(X < t \leq Y))dt \\
&= \int_{-\infty}^{\infty} H(P(X \leq t) + P(Y \leq t) - 2P(X \leq t, Y \leq t)))dt \\
&\geq \int_{-\infty}^{\infty} H(F_X(t) + F_Y(t) - 2\min(F_X(t), F_Y(t)))dt \\
&= \int_{-\infty}^{\infty} H(|F_X(t) - F_Y(t)|)dt = \boldsymbol{\theta}_H(X,Y)
\end{aligned}$$

Minimal and maximal distances

- In fact, the Fréchet-Hoeffding inequality is not unrelated to the Cambanis-Simons-Stout result.
- The minimal and the maximal functionals are obtained at the upper and the lower Fréchet-Hoeffding bounds and they can also be represented in terms of random variables as

$$\hat{\mu}_\phi(X, Y) = E\phi(F_X^{-1}(U), F_Y^{-1}(U)) \tag{27}$$

and

$$\check{\mu}_\phi(X, Y) = E\phi(F_X^{-1}(U), F_Y^{-1}(1 - U)) \tag{28}$$

where U is a uniformly distributed random variable in the interval $(0, 1)$.

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.

Rachev, S.T.
Probability Metrics and the Stability of Stochastic Models
Wiley, Chichester, UK, 1991.