Monge Kantorovich Problem and the Theory of Probability Metrics

Prof. Dr. Svetlozar (Zari) Rachev

Frey Family Foundation Chair-Professor,
Applied Mathematics and Statistics, Stony Brook University
Chief Scientist, FinAnalytica

Spring 2012

Outline

- Monge-Kantorovich Mass Transportation Problem
- Duality Theorems and Kantorovich Problem
- Kemperman problem
- MKP and Probability Metrics
- Ideal metrics in Probability Theory
- Applications and Open Problems
- References

Monge-Kantorovich Problem

- There are six known versions of the Monge-Kantorovich problem (MKP).
- We will discuss three:

  I. Monge transportation problem,

  II. Kantorovich's mass transference problem,

  III. Kantorovich-Rubinstein-Kemperman problem of multistaged shipping.

Monge Problem

Gaspard Monge, Comte de Péluse (9 May 1746[1] – 28 July 1818) was a French mathematician, revolutionary, and was inventor of descriptive geometry. During the French Revolution, he was involved in the complete reorganization of the educational system, founding the École Polytechnique. He also served as minister of the Marine during the revolution.

(I) Monge transportation problem

In 1781, Monge formulated the following problem in studying the most efficient way of transporting soil.

Split two equally large volumes into infinitely small particles and then associate them with each other so that the sum of the products of these paths of the particles to a volume is least. Along what paths must the particles be transported and what is the smallest transportation cost?

(I) Monge transportation problem

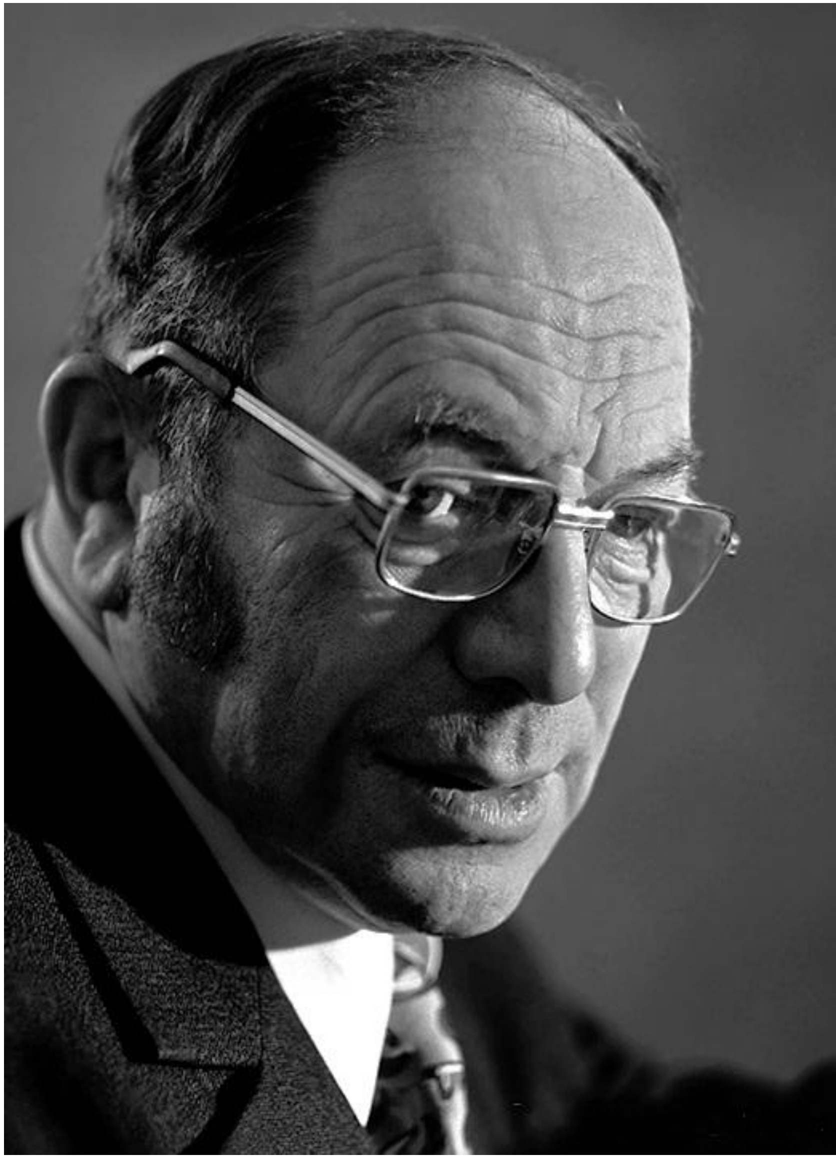
- In other words, two sets S_1 and S_2 are the supports of two masses μ_1, μ_2 with equal total weight

$$\mu_1(S_1) = \mu_2(S_2)$$

- The 'initial' mass μ_1 is to be transported from S_1 to S_2 so that the result is the 'final' mass μ_2.
- The transportation should be realized in such a way as to minimize the total labor involved.

Kantorovich Problem

Leonid Vitaliyevich Kantorovich (Russian: Леони́д Вита́льевич Канторо́вич) (19 January 1912 – 7 April 1986) was a Soviet mathematician and economist, known for his theory and development of techniques for the optimal allocation of resources. He was the winner of the Nobel Prize in Economics in 1975 and the only winner of this prize from the USSR.



(II) Monge – Kantorovich problem

In the Monge problem

- Let A and B be initial and final volumes.
- For any set $a \subset A$ and $b \subset B$ let $P(a,b)$ be the fraction of volume of A that was transferred from a into b.
- Note: $P(a,B)$ is equal to the ratio of volumes of a and A and $P(A,b)$ is equal to the ratio of volumes of b and B, respectively.

(II) Kantorovich's mass transference problem

- In general we need not assume that A and B are of equal volumes; rather they are bodies with equal masses though not necessarily uniform densities.
- Let $P_1(\cdot)$ and $P_2(\cdot)$ be the probability measures on the space U, respectively describing the masses of A and B.
- A shipping plan would be a probability measure P on $U \times U$ such that its projections on the first and second coordinates are P_1 and P_2, respectively.
- The amount of mass shipped from a neighborhood dx of x into the neighborhood dy of y is then proportional to $P(dx, dy)$.
- If the unit cost of shipment from x to y is $c(x, y)$ then the total cost is

$$\int_{U\times U} c(x,y)P(dx,dy) \tag{1}$$

(II) Kantorovich's mass transference problem

Thus we see that minimization of transportation costs can be formulated in terms of finding a distribution of $U \times U$ whose marginals are fixed, and such that the double integral of the cost function is minimal. This is the so-called *Kantorovich formulation of the Monge problem* which in abstract form is as follow:

Suppose that P_1 and P_2 are two Borel probability measures given on a separable metric space (s.m.s.) (U, d) and $\mathcal{P}^{(P_1,P_2)}$ is the space of all Borel probability measures P on $U \times U$ with fixed marginals $P_1(\cdot) = P(\cdot \times U)$ and $P_2(\cdot) = P(\cdot \times U)$. Evaluate the functional

$$\mathcal{A}_c(P_1, P_2) = inf\left\{\int_{U\times U} c(x,y)P(dx,dy) : P \in \mathcal{P}^{(P_1,P_2)}\right\} \qquad (2)$$

Where $c(x, y)$ is a given continuous non-negative function on $U \times U$.

(II) Kantorovich problem

- We shall call the functional (2) *Kantorovich's functional*, (resp. Kantorovich metric) if $c = d$.
- The measures P_1 and P_2 may be viewed as the initial and final distribution of mass and $\mathscr{P}^{(P_1,P_2)}$ as the space of admissible transference plans.
- If the infimum in (2) is realized for some measure $P^* \in \mathscr{P}^{(P_1,P_2)}$ then P^* is said to be the *optimal transference plan*.
- The function $c(x, y)$ can be interpreted as the cost of transferring the mass from x to y.

(II) Kantorovich’s mass transference problem

Remark 1

- Kantorovich’s formulation differs from the Monge problem in that the class $\wp^{(P_1,P_2)}$ is broader than the class of one-to-one transference plans in Monge’s sense.
- Sudakov (1976) showed that if measures P_1 and P_2 are given on a bounded subset of a finite-dimensional Banach space and are absolutely continuous with respect to Lebesgue measure, then there exists an optimal one-to-one transference plan.

Remark 2

- Another example of the MKP: Assigning army recruits to jobs to be filled.
  - The flock of recruits has a certain distribution of parameters such as education, previous training, and physical conditions.
  - The distribution of parameters which are necessary to fill all the jobs might not necessarily coincide with one of the contingents.
  - There is a certain cost involved in training of an individual for a specific job depending on the job requirements and individual parameters, thus the problem of assigning recruits to the job and training them so that the total cost is minimal can be viewed as a particular case of the MKP.

(II) Kantorovich's mass transference problem

- Comparing the definition of $\mathcal{A}_c(\mathrm{P}_1, \mathrm{P}_2)$ with the definition of minimal distance $\hat{\mu}$ we see that
$$\mathcal{A}_c = \hat{\mu} \tag{3}$$
- For any compound distance μ of the form
$$\mu(P) = \mu_c(P) = \int_{U \times U} c(x, y) P(dx, dy) \qquad P \in \mathcal{P}_2 \tag{4}$$
(Recall that $\mathcal{P}_k$ is the set of all Borel probability measures on the Cartesian product U^k.)
- If $\mu(P) = \mathcal{L}_H(P) := \int H(d(x, y)) P(dx, dy), H \in \mathcal{H}, P \in \mathcal{P}_2$ is the H-average compound distance, then $\mathcal{A}_c = \hat{\mathcal{L}}_H$.
- This example seems to be the most important for the point of view of the theory of probability metrics.
- For this reason we will devote special attention to the mass transportation problem with cost function $c(x, y) = H(d(x, y))$.

Kantorovich and Rubinstein (1957)

- Studied the problem of transferring masses in cases where transits are permitted.
- Rather than shipping a mass from a certain subset of U to another subset of U in one step, the shipment is made in n stages. Namely, we ship $A=A_1$ to volume A_2, then A_2 to $A_3, \dots, A_{n-1}$ to $A_n = B$.
- Let $\gamma_n(a_1, a_2, a_3, \dots, a_n)$ be measure equal to the total mass which was removed from the set a_1 and on its way to a_n passed the sets $a_2, a_3, \dots, a_{n-1}$.
- If $c(x, y)$ is the unit cost o transportation from x to y then the total cost under such a transportation plan is

$$\int_{U\times U} c(x,y)\gamma_n(dx \times dy \times U^{n-2}) + \sum_{i=2}^{n-2} \int_{U\times U} c(x,y)\gamma_n\left(U^{i-1} \times dx \times dy \times U^{n-i-1}\right)$$
$$+ \int_{U\times U} c(x,y)\gamma_n(U^{n-2} \times dx \times dy) =: \int_{U\times U} c(x,y)\Gamma_n(dx \times dy) \qquad (5)$$

- **George Bernard Dantzig** (November 8, 1914 – May 13, 2005) was an American mathematical scientist who made important contributions to operations research, computer science, economics, and statistics.
- Dantzig is known for his development of the simplex algorithm, an algorithm for solving linear programming problems, and his work with linear programming, some years after it was invented by the Soviet mathematician & economist Leonid Kantorovich.[1] In statistics, Dantzig solved two open problems in statistical theory, which he had mistaken for homework after arriving late to a lecture of Jerzy Neyman.[2]
- Dantzig was the Professor Emeritus of Transportation Sciences and Professor of Operations Research and of Computer Science at Stanford.

Kemperman(1983)

- A more sophisticated plan
- Consists of a sequence of transportation subplans γ_n, $n = 2,3,\ldots$
- Each subplan γ_n need not transfer the whole mass from A to B, rather only a certain part of it. However, combined they complete the transshipment of mass, that is,

$$P_1(A) = \sum_{n=2}^{\infty} \gamma_n(A \times U^{n-1}) \tag{6}$$

and

$$P_2(B) = \sum_{n=2}^{\infty} \gamma_n(U^{n-1} \times B) \tag{7}$$

- The total cost of transshipment under this sequential transportation plan will be the sum of costs of each subplan and is equal to

$$\int_{U\times U} c(x,y)Q(dx\times dy) \tag{8}$$

where

$$Q(A\times B) = \sum_{n=2}^{\infty} \Gamma_n(A\times B) \tag{9}$$

and Γ_n is defined by (5),

$$\Gamma_n(A\times B) \coloneqq \gamma_n(A\times B\times U^{n-2}) + \sum_{i=2}^{n-2} \gamma_n\left(U^{i-1} + A\times B\times U^{n-i-1}\right) + \gamma_n(U^{n-2}\times A\times B)$$

Note

- Q is not necessarily a probability measure. The marginals of Q are equal to

$$Q_1(A) = \sum_{n=2}^{\infty}\left(\gamma_n(A \times U^{n-1}) + \sum_{i=1}^{n-2}\gamma_n\left(U^{i-1} \times A \times U^{n-i-1}\right)\right) \quad (10)$$

and

$$Q_2(B) = \sum_{n=2}^{\infty}\left(\gamma_n(U^{n-1} \times B) + \sum_{i=1}^{n-2}\gamma_n\left(U^{i-1} \times B \times U^{n-i-1}\right)\right) \quad (11)$$

respectively, combining equalities (6) (7) and (10) (11), we obtain

$$Q_1(A) - P_1(A) = Q_2(A) - P_2(A) = \sum_{n=3}^{\infty}\sum_{i=1}^{n-2}\gamma_n\left(U^i \times A \times U^{n-i-1}\right) \quad (12)$$

for any $A \in \mathcal{B}(U)$.

- Denote the space of all translocations of masses (without transits permitted) by $\mathscr{P}^{(P_1,P_2)}$ (cf. (2)). Under the *translocations of masses with transits permitted* we will understand the finite Borel Measure Q on $\mathcal{B}(U \times U)$ such that
$$Q(A \times U) - Q(U \times A) = P_1(A) - P_2(A) \tag{13}$$
for any $A \in \mathcal{B}(U)$.
- Denote the space of all Q satisfying (13) by $\mathcal{L}^{(P_1,P_2)}$. Let a continuous non-negative function $c(x,y)$ be given that represents that cost of transferring a unit mas from x to y. The total cost of transferring the given mass distributions P_1 and P_2 is given by
$$\mu_c(P) := \int_{U \times U} c(x,y) P(dx,dy) \quad if\ P \in \mathscr{P}^{(P_1,P_2)} \tag{14}$$
(cf. (2)) or
$$\mu_c(Q) := \int_{U \times U} c(x,y) Q(dx,dy) \quad if\ Q \in \mathcal{L}^{(P_1,P_2)} \tag{15}$$

- Hence, if μ_c is a p. distance then the minimal distance

$$\hat{\mu}_c(P_1, P_2) = \inf\{\mu_c(P) : P \in \mathcal{P}^{(P_1,P_2)}\} \quad (16)$$

may be viewed as the minimal translocation cost while the minimal norm

$$\mathring{\mu}_c(P_1, P_2) = \inf\{\mu_c(Q) : Q \in \mathcal{L}^{(P_1,P_2)}\} \quad (17)$$

may be viewed as the minimal translocation cost in case of transits permitted.

- The problem of calculating of exact value of $\hat{\mu}_c$(for general c) is known as the *Kantorovich problem* and $\hat{\mu}_c$ is called the *Kantorovich functional* (see Equality (2)).
- Similarly, the problem of evaluating $\dot{\mu}_c$ is known as the *Kantorovich-Rubinstein problem* and $\dot{\mu}_c$ is said to be the *Kantorovich-Rubinstein functional*. Some authors refer to $\dot{\mu}_c$ as the *Wasserstein norm* if $c = d$.

MKP and Probability Metrics

The origin of The Theory of Probability Metrics and Stability of Stochastic models

Kantorovich Metric

Kolmogorov Metric

Levy Metric

Prokhorov Metric

- Rachev, S. T., Rueschendorf, L., *Mass Transportation Problems, Vol II: Applications*, Springer, New York, 1999
- Rachev, S. T., Rueschendorf, L., *Mass Transportation Problems, Vol I: Theory*, Springer, New York, 1998
- Rachev, S. T., *Probability Metrics and the Stability of Stochastic Models*, Wiley, Chichester, New York, 1991

Kolmogorov Metric

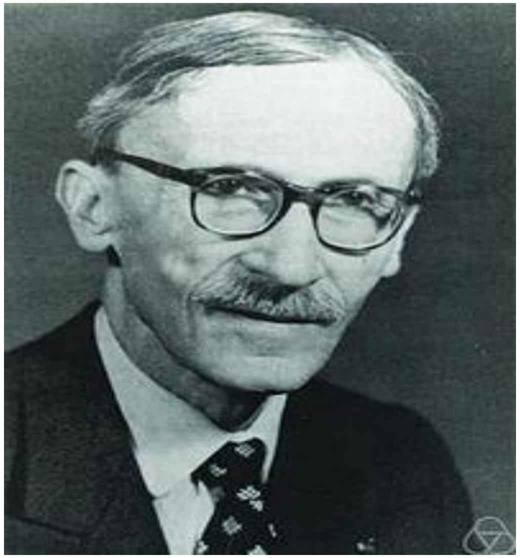
Kolmogorov (Uniform) Metric is the uniform distance between two distribution functions

Andrey Nikolayevich Kolmogorov, (born April 25 [April 12, Old Style], 1903, Tambov, Russia—died Oct. 20, 1987, Moscow), Russian mathematician whose work influenced many branches of modern mathematics, especially harmonic analysis, probability, set theory, information theory, and number theory.

The Kolmogorov–Smirnov statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples.

Levy Metric

Kolmogorov's metric generates the uniform convergence in the space of probability distributions. The Levy metric is topologically weaker, it generates the weak convergence of probability distributions. Intuitively, if between the graphs of F and G one inscribes squares with sides parallel to the coordinate axes (at points of discontinuity of a graph vertical segments are added), then the side-length of the largest such square is equal to $L(F, G)$, it is Hausdorff distance between the completed graphs of two distribution functions.



Paul Pierre Lévy (15 September 1886 – 15 December 1971) was a Jewish French mathematician who was active especially in probability theory, introducing martingales and Lévy flights. Lévy processes, Lévy measures, Lévy's constant, the Lévy distribution, the Lévy skew alpha-stable distribution, the Lévy area, the Lévy arcsine law, and the fractal Lévy C curve are also named after him.

Prokhorov Distance = Levy-Prokhorov Metric

Yuri Vasilevich Prokhorov
Юрий Васильевич Прохоров

Born	1929-12-15 Moscow, USSR
Institutions	Russian Academy of Sciences
Doctoral advisor	Andrey Nikolaevich Kolmogorov
	Lenin Prize (1970) Order of the Red Banner of Labour (1975, 1979) Prokhorov Metric is the solution of MKP when the objective function is the Ky Fan Metric = distance in probability. It generates the weak convergence for probability measures in separable metric spaces

Dudley Metric

- Prokhorov metric is topologically equivalent to Dudley Metrics, which is the solution of MKP with cost function $c(x,y) = \min(1,d(x,y))$, on a separable metric space (U,d).

- **Richard Mansfield Dudley** is Professor of Mathematics at the Massachusetts Institute of Technology. He received his PhD at Princeton University in 1962 under the supervision of Edward Nelson and Gilbert Hunt. He was a Putnam Fellow in 1958.
- He has published over a hundred papers in peer-reviewed journals and has written several books. His specialty is probability theory and statistics, especially empirical processes.

Fortet-Mourier Metric – generates weak convergence plus convergence of the moments – solution of MKP for special cost function c(x,y)

- The functional $\dot{\mu}_c$ is frequently used in mathematical-economical models (cf. for example, Bazaraa and Jarvis, 1971, Chapter 9), but it is not applied in probability theory.
- Observe, however, the following relationship between the Fortet-Mourier metric

$$\zeta(P,Q;\mathcal{G}^P) = \sup\{\int_U f\,\mathrm{d}(P-Q): f: U \to \mathbb{R}, \text{and}$$
$$|f(x)-f(y)| \le d(x,y)\max[1, d(x,a)^{p-1}, d(y,a)^{p-1}] \quad \forall x,y \in U$$

and the minimal norm $\dot{\mu}_c$:

$$\zeta(P,Q;\mathcal{G}_p) = \dot{\mu}_c(P,Q)$$

where the cost function is given by

$$c(x,y) = d(x,y)\max[1, d(x,a)^{p-1}, d^{p-1}(y,a)], p \ge 1$$

Kolmogorov and Zolotarev Ideal Metrics for Sums of Independent Random Variables

- We will introduce two ideal metrics of convolution type on the space $\mathfrak{X}$.
- These ideal metrics will be used to provide exact convergence rates for convergence to an α-stable random variable in the Banach space setting.
- Moreover, the rates will hold with respect to a variety of uniform metrics on $\mathfrak{X}$.

Ideal Metrics for Sums of Independent Random Variables

- Let $(U, \|\cdot\|)$ be a complete separable Banach space equipped with the usual algebra of Borel sets $\mathcal{B}(U)$
- Let $\mathfrak{X} := \mathfrak{X}(U)$ be the vector space of all random variables defined on a probability space $(\Omega, \mathcal{A}, \Pr)$ and taking values in U.
- We will choose to work with simple probability metrics on the space $\mathfrak{X}$ instead of the space $\mathcal{P}(U)$.
- We will show that certain 'convolution' metrics on $\mathfrak{X}$ may be used to provide exact rates of convergence of normalized sums to a stable limit law. They will play the role of 'ideal metrics' for the approximation problems under consideration.

Ideal Metrics for Sums of Independent Random Variables

- 'Traditional' metrics for the rate of convergence in the CLT are uniform type metrics.
- Having exact estimates in terms of the 'ideal' metrics we shall pass to the uniform estimates by using the Bergstrom convolution method.
- The rates of convergence, which hold uniformly in n will be expressed in terms of a variety of uniform metrics on $\mathfrak{X}$.

Ideal Metrics for Sums of Independent Random Variables: Definition

Definition (Zolotarev). A p. semimetric $\mu\colon \mathfrak{X} \times \mathfrak{X} \to [0, \infty]$ is called an *ideal (probability) metric* of order $r \in \mathbb{R}$ if for any random variables $X_1, X_2, Z \in \mathfrak{X}$ and any non-zero constant c the following two properties are satisfied

(i) *Regularity:* $\mu(X_1 + Z, X_2 + Z) \leq \mu(X_1 X_2)$, and

(ii) *Homogeneity of order r:* $\mu(cX_1, cX_2) = |c|^r \mu(X_1, X_2)$.

- When μ is a simple metric, i.e., its values are determined by the marginal distributions of the random variables being compared, then it is assumed in addition that the random variable Z is independent of X_1 and X_2 in condition (i).
- All metrics μ in this section are simple.

MKP and Probability Metrics

- Rachev S. T., *The Monge-Kantorovich mass transference problem and its stochastic applications* Theor. Probab. Appl., Vol. 29, No. 4, 1985, 647-676

Rachev S.T **Ph.D. in Mathematics**, Lomonosov University (Moscow), Faculty of Mechanics and Mathematics, October 12, 1979, **Dissertation**: "*The structure of the metrics in the space of random variables and their distributions*.

Rachev S.T. " **Doctor of Science (Habilitation)** in Physics and Mathematics, Steklov Mathematical Institute, Moscow, April 10, 1986. **Dissertation**: "*Probability metrics and their applications to the stability problems for stochastic models*"

- S. T. Rachev, *On a problem of Dudley.* Soviet Math. Dokl. 1984, Vol 29, No. 2, 162-164 (Presented by A.N. Kolmogorov)
- Rachev S. T., *On the ζ-structure of the average and uniform distances.* Soviet Math. Dokl., Vol. 30, 1984, No.2, 369-372 (Presented by A.N. Kolmogorov)
- Rachev S. T., Maejima M.; *An ideal metric and the rate of convergence to a self-similar process*, Ann. Probability, Vol. 15, 1987, 702-707.

Remark 1

- Zolotarev (1976 b,d) showed the existence of an ideal metric of a given order $r \geq 0$ and he defined the ideal metric

$$\zeta_r(X_1, X_2) := \sup\{|\mathbb{E}(f(X_1) - f(X_2))| : |f^{(m)}(x) - f^{(m)}(y)| \leq \|x - y\|^\beta\} \qquad (1)$$

where

  - $m = 0,1,\dots$ and $\beta \in (0,1]$ satisfy $m + \beta = r$.
  - $f^{(m)}$ denotes the mth Frechet derivative of f for $m \geq 0$ and $f^{(0)}(x) = f(x)$.

- He also obtained an upper bound for ζ_k (k integer) in terms of the *difference pseudomoment* $\boldsymbol{\kappa}_k$ where for $r > 0$

$$\boldsymbol{\kappa}_r(X_1, X_2) := \sup\{|\mathbb{E}(f(X_1) - f(X_2))| : |f(x) - f(y)| \leq \|x\|x\|^{r-1} - y\|y\|^{r-1}\|\}$$

- If $U = \mathbb{R}$, $\|x\| = |x|$, then

$$\boldsymbol{\kappa}_r(X_1, X_2) := r \int |x|^{r-1} |F_{X_1}(x) - F_{X_2}(x)|\, dx \qquad r > 0 \qquad (2)$$

where F_X denotes the distribution function for X.

Remark 2

- For each $X_1, X_2 \in \mathfrak{X}$ we write $X_1 + X_2$ to mean the sum of independent random variable with laws Pr_{X_1} and Pr_{X_1} respectively.
- For any $X \in \mathfrak{X}$, p_X denotes the density of X if it exists.
- We reserve the letter Y_α (or Y) to denote a *symmetric stable random variable* with parameter $\alpha \in (0,2]$, i.e. $Y_\alpha \stackrel{\text{def}}{=} -Y_\alpha$
- For any $n = 1,2, \dots, X_1' + \cdots + X_n' \stackrel{\text{def}}{=} n^{1/\alpha} Y_\alpha$, where $X_1', X_2', \dots, X_n'$ are i.i.d. random variables with the same distribution as Y_α.
- If $Y_\alpha \in \mathfrak{X}(\mathbb{R})$ we assume that Y_α has characteristic function
$$\phi_Y(t) = exp\{-|t|^\alpha\} \qquad t \in \mathbb{R}$$
- For any $f: U \to \mathbb{R}$,
$$\|f\|_L := sup_{x \neq y} \frac{|f(x) - f(y)|}{\|x - y\|}$$
denotes the Lipschitz norm of f, $\|f\|_\infty$ the essential supremum of f, and when $U = \mathbb{R}^k$, $\|f\|_p$ denotes the L^p norm,
$$\|f\|_p^p := \int_{\mathbb{R}^k} |f(x)|^p dx, \qquad p \geq 1$$

- Letting $X, X_1, X_2, \dots$ denote i.i.d. random variables and Y_α denote an α-stable random variable we shall use ideal metrics to describe the rate of convergence

$$\frac{X_1 + \cdots + X_n}{n^{1/\alpha}} \xrightarrow{w} Y_\alpha \qquad (3)$$

with respect to the following uniform metrics on $\mathfrak{X}$, ($\xrightarrow{w}$ stands for the weak convergence).

- *Total variation metrics*

$$\sigma(X_1, X_2) := \sup_{A\in\mathcal{B}(U)}|\Pr\{X_1 \in A\} - \Pr\{X_2 \in A\}|$$
$$:= \sup\{|\mathbb{E}f(X_1) - \mathbb{E}f(X_2)|: f: U \to \mathbb{R} \text{ is measurable and for any } x, y \in B, |f(x) - f(y)| \le \mathbb{I}(x, y) \text{ where } \mathbb{I}(x, y) = 1 \; if \; x \neq y \text{ and } 0 \text{ otherwise}\}, X_1, X_2 \in \mathfrak{X}(U) \quad (4)$$

and

$$\mathrm{Var}(X_1, X_2) := \sup\{|\mathbb{E}f(X_1) - \mathbb{E}f(X_2)|: f: U \to \mathbb{R} \text{ is measurable and } \|f\|_\infty \le 1\} = 2\sigma(X_1, X_2), X_1, X_2 \in \mathfrak{X}(U) \quad (5)$$

In $\mathfrak{X}(\mathbb{R}^n)$ we have

$$\mathrm{Var}(X_1, X_2) := \int \left| d(F_{X_1} - F_{X_2}) \right|.$$

- *Uniform metric between densities*

  (p_X denotes the density for $X \in \mathfrak{X}(\mathbb{R}^k)$)

  $$\ell(X_1, X_2) := \text{ess sup}_x \left| p_{X_1}(x) - p_{X_2}(x) \right| \qquad (6)$$

- *Uniform metric between characteristic functions*

  $$\chi(X_1, X_2) := \sup_{t \in \mathbb{R}} \left| \phi_{X_1}(t) - \phi_{X_2}(t) \right| \quad X_1, X_2 \in \mathfrak{X}(\mathbb{R}) \quad (7)$$

  where ϕ_X denotes the characteristic function of X.

  The metric χ is topologically weaker than Var, which is itself topologically weaker than ℓ by Scheffe's theorem, see Billingsley (1968, p. 224)

- *Kolmogorov metric*

$$\rho(X_1, X_2) := \sup_{x\in\mathbb{R}} \left|F_{X_1}(x) - F_{X_2}(x)\right| \tag{8}$$

- *Weighted χ-metric*

$$\chi_r(X_1, X_2) := \sup_{t\in\mathbb{R}} |t|^{-r} \left|\phi_{X_1}(t) - \phi_{X_2}(t)\right| \tag{9}$$

- *L^p-version of ζ_m*

$$\zeta_{\mathrm{m,p}}(X_1, X_2) := \sup\left\{|\mathbb{E}f(X_1) - f(X_2)| : \left\|f^{(m+1)}\right\|_q \le 1\right\}$$

$$\frac{1}{p} + \frac{1}{q} = 1 \quad m = 0,1,2,\ldots \tag{10}$$

If $\zeta_{\mathrm{m,p}}(X_1, X_2) < \infty$ then

$$\zeta_{\mathrm{m,p}}(X_1, X_2) = \left\|\int_{-\infty}^{\bullet} \frac{(\bullet - t)^m}{m!} d(F_{X_1}(t) - F_{X_2}(t))\right\|_p$$

see Kalashnikov and Rachev (1988).

- *Kantorovich ℓ_p-metric:*

$$\ell_p^p(X_1, X_2) := \sup\left\{\int f dF_{X_1} + \int g dF_{X_2} : \|f\|_\infty + \|f\|_L \leq \infty \right.$$

$$\|g\|_\infty + \|g\|_L \leq \infty, f(x) + g(y) \leq \|x - y\|^p$$

$$\left. \forall x, y \in \mathbb{R}\right\}, p \geq 1 \tag{11}$$

- Now we define the ideal metrics of order $r-1$ and r, respectively.
- Let $\theta \in \mathfrak{X}(\mathbb{R}^k)$, $\theta \stackrel{\text{def}}{=} -\theta$ and define for every $r > 0$ the *convolution* (probability) metric

$$\boldsymbol{\mu}_{\theta,r}(X_1, X_2) := \sup_{h\in\mathbb{R}} |h|^r \ell(X_1 + h\theta, X_2 + h\theta) \quad X_1, X_2 \in \mathfrak{X}(\mathbb{R}^k) \quad (12)$$

- Thus, each random variable θ generates a metric $\boldsymbol{\mu}_{\theta,r}$, $r > 0$. When $\theta \in \mathfrak{X}(U)$ we will also consider convolution metrics of the form

$$\mathbf{v}_{\theta,r}(X_1, X_2) := \sup_{h\in\mathbb{R}} |h|^r \mathrm{Var}(X_1 + h\theta, X_2 + h\theta) \quad X_1, X_2 \in \mathfrak{X}(U) \quad (13)$$

Ideal Metrics for Sums of Independent Random Variables: Lemmas

Lemma 1

- For all $\theta \in \mathfrak{X}$ and $r > 0$, $\boldsymbol{\mu}_{\theta,r}$ is an ideal metric of order $r - 1$.

Lemma 2

- For all $\theta \in \mathfrak{X}$ and $r > 0$, $\boldsymbol{\nu}_{\theta,r}$ is an ideal metric of order r.

- In general, $\boldsymbol{\mu}_{\theta,r}$ and $\boldsymbol{\nu}_{\theta,r}$ are actually only semimetrics, but this distinction is not of importance in what follows and so we omit it.
- When θ is a symmetric α-stable random variable we will write $\boldsymbol{\mu}_{\alpha,r}$ and $\boldsymbol{\nu}_{\alpha,r}$ or simply $\boldsymbol{\mu}_r$ and $\boldsymbol{\nu}_r$ when it is understood, in place of $\boldsymbol{\mu}_{\theta,r}$ and $\boldsymbol{\nu}_{\theta,r}$.
- We will describe the special properties of the ideal convolution (or smoothing) metrics $\boldsymbol{\mu}_{\theta,r}$ and $\boldsymbol{\nu}_{\theta,r}$. We first verify ideality.

We now show that both $\boldsymbol{\mu}_{\theta,r}$ and $\boldsymbol{\nu}_{\theta,r}$ are bounded above by the difference pseudomoment whenever θ has a density which is smooth enough.

Lemma 3

- Let $k \in \mathbb{N}^+ := \{0,1,2,\dots\}$ and suppose that $X, Y \in \mathfrak{X}(\mathbb{R})$ satisfy $\mathbb{E}X^j = \mathbb{E}Y^j, j = 1,\dots,k-2$. Then for every $\theta \in \mathfrak{X}(\mathbb{R})$ with a density g which is $k-1$ times differentiable

$$\boldsymbol{\mu}_{\theta,k}(X_1, X_2) \leq \frac{\left\| g^{(k-1)} \right\|_\infty}{(k-1)!} \boldsymbol{\kappa}_{k-1}(X_1, X_2) \qquad (14)$$

Lemma 4

- For every $\theta \in \mathfrak{X}(\mathbb{R})$ with a density g which is m times differentiable and for all $X_1, X_2 \in \mathfrak{X}(\mathbb{R})$,
$$\boldsymbol{\mu}_{\theta,r}(X_1, X_2) \leq C(m,p,g)\zeta_{m-1,p}(X_1, X_2) \tag{19}$$
where $r = m + 1/p$, $m \in \mathbb{N}^+$, and
$$C(m,p,g) := \left\|g^{(m)}\right\|_q \quad \frac{1}{p} + \frac{1}{q} = 1 \tag{20}$$

Lemma 5

- Under the hypotheses of Lemma 4 we have
$$\boldsymbol{\nu}_{\theta,r}(X_1, X_2) \leq C(r,g)\zeta_r(X_1, X_2) \tag{21}$$
where $C(r,g)$ is a finite constant, $r \in \mathbb{N}^+$.

Lemma 6

- Let $m \in \mathbb{N}^+$ and suppose $\mathbb{E}\left(X_1^j - X_2^j\right) = 0, j = 0,1,\dots,m$. Then for $p \in [1,\infty)$

$$\zeta_{m,p}(X_1,X_2) \le \begin{cases} \boldsymbol{\kappa}_1^{1/p}(X_1,X_2) & \text{if } m = 0 \\ \dfrac{\Gamma(1+1/p)}{\Gamma(r)}\boldsymbol{\kappa}_r(X_1,X_2) & \text{if } m = 1,2,\dots, r = m + 1/p \end{cases}$$

- Also, for $r = m + 1/p$

$$\zeta_{m,p}(X_1,X_2) \le \zeta_r(X_1,X_2)$$

Ideal Metrics for Sums of Independent Random Variables: Lemmas

- Lemmas 4, 5 and 6 describe the conditions under which $\zeta_{\theta,r}$ (resp. $\mathbf{v}_{\theta,r}$) is finite. Thus by (19) and (22) we have that for $r > 1$

$$\begin{cases} \mathbb{E}\left(X_1^j - X_2^j\right) = 0, \; j = 0,1,\dots,m-1 \\ r := m + 1/p \\ \boldsymbol{\kappa}_{r-1}(X_1, X_2) < \infty \end{cases} \Rightarrow \boldsymbol{\mu}_{\theta,r}(X_1, X_2) < \infty \qquad (23)$$

- For any θ with density g such that $\left\|g^{(m-1)}\right\|_q \leq \infty, \frac{1}{p} + \frac{1}{q} = 1$. In particular, if θ is an α-stable, then

$$\begin{cases} \int x^j d\left(F_{X_1} - F_{X_2}\right)(x) = 0, \; j = 0,1,\dots,m-1 \\ r := m + 1/p \\ \boldsymbol{\kappa}_{r-1}(X_1, X_2) < \infty \end{cases} \Rightarrow \boldsymbol{\mu}_{\alpha,r}(X_1, X_2) < \infty \qquad (24)$$

- Similarly,

$$\begin{cases} \int x^j d\left(F_{X_1} - F_{X_2}\right)(x) = 0, \; j = 0,1,\dots,r-1 \\ r \in \mathbb{N}^+ \\ \boldsymbol{\kappa}_{r}(X_1, X_2) < \infty \end{cases} \Rightarrow \mathbf{v}_{\alpha,r}(X_1, X_2) < \infty \qquad (25)$$

- We conclude our discussion of the ideal metrics $\boldsymbol{\mu}_{\alpha,r}$ and $\mathbf{v}_{\alpha,r}$ by showing that they satisfy the same weak convergence properties as do the Kantorovich distance ℓ_p and the pseudomoments $\boldsymbol{\kappa}_r$.

Theorem

- Let $k \in \mathbb{N}^+$, $0 < \alpha \leq 2$, and $X_n, U \in \mathfrak{X}(\mathbb{R})$ with $\mathbb{E}X_n^j = \mathbb{E}U^j$, $j = 1, \dots, k-2$ and $\mathbb{E}|X_n|^{k-1} + \mathbb{E}|U|^{k-1} < \infty$. If k is odd then the following are equivalent as $n \to \infty$:
  (i) $\boldsymbol{\mu}_{\alpha,k}(X_n, U) \to 0$
  (ii) (a) $X_n \xrightarrow{w} U$ and (b) $\mathbb{E}|X_n|^{k-1} \to \mathbb{E}|U|^{k-1}$
  (iii) $\ell_{k-1}(X_n, U) \to 0$
  (iv) $\boldsymbol{\kappa}_{k-1}(X_n, U) \to 0$
  (v) $\boldsymbol{\nu}_{\alpha,k-1}(X_n, U) \to 0$

Claim

- Let $0 < \alpha \leq 2$ and consider the associated metric $\boldsymbol{\mu}_r := \boldsymbol{\mu}_{r,\alpha}$. For all k there is a constant $\beta := \beta(\alpha, k) < \infty$ such that for all $X, U \in \mathfrak{X}(\mathbb{R})$

$$\boldsymbol{\mu}_k(X,U) \geq \beta \left| \int_{\mathbb{R}} F_X^{(2-k)}(z) - F_U^{(2-k)}(z) dz \right| \qquad (26)$$

Here $F^{(2-k)}$ is as in $F_X^{-k}(x) := \int_{-\infty}^{x} \frac{(x-t)^k}{k!} dF_X(t)$

Seminars on Stability Problems for Stochastic Models have a long tradition. They were started by Prof. V.M. Zolotarev in the 1970's.

- Limit theorems and stability
- Asymptotic for stochastic processes
- Stable distributions and processes
- Topics in mathematical statistics
- Queuing theory and modeling of communication systems
- Discrete probabilistic models
- Characterization of probability distributions
- Insurance and financial mathematics
- Generalized stability and convolution

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

Open Problems

1. Monge-Kantorovich Problem – explicit solutions.
2. Compound ideal metrics of order r>1 - Do they exist?
3. Applications of Monge Kantorovich problems in stochastic ordering, optimization, and others.

See references next...

References

1) Rachev, S. T., Klebanov, Lev., Stoyanov, S., Fabozzi, F., The Method of Distances in Probability, Statistics and Applications, Springer, NY, 2012, Forthcoming
2) Rachev, S. T., Stoyanov, S. V., Fabozzi, F., A Probability Metrics Approach to Financial Risk Measures, Wiley . Blackwell, 2011
3) Rachev, S. T., Stoyanov, S., Fabozzi, F., Advanced Stochastic Models, Risk Assessment and Portfolio Optimization: The Ideal Risk, Uncertainty, and Performance Measures, John Wiley, Finance, 2007
4) Rachev, S. T., Rueschendorf, L., Mass Transportation Problems, Vol II: Applications, Springer, New York, 1999
5) Rachev, S. T., Rueschendorf, L., Mass Transportation Problems, Vol I: Theory, Springer, New York, 1998
6) Rachev, S. T., Probability Metrics and the Stability of Stochastic Models, Wiley, Chichester, New York, 1991
7) Differential geometry – see Germaine (1886), Appell (1884, 1928), Vygodskii (1936), Sudakov (1976);
8) Functional analysis – Kantorovich and Akilov (1984), Levin and Milyutin (1979)

References

9) Linear programming – Hoffman (1961), Berge and Ghouila-Houri (1965, Chap.9), Vershik (1970), Rubinstein (1970), Kemperman (1983), Knott and Smith (1984), Smith and Knott (1987), Barnes and Hoffman (1985), Anderson and Nash (1987), Kellerer (1988)
10) Probability theory – Kantorovich and Rubinstein (1958), Fernique (1981), Dudley (1976, 1989), Zolotarev (1983b), Kellerer (1984a,b), Ruschendorf (1985b)
11) Mathematical statistics – Huber (1981), Bickel and Freedman (1981)
12) Information theory and cybernetics – Wasserstein (1969), Gray et al. (1975), Gray and Ornstein (1979), Gray et al. (1980)
13) Statistical physics – Dobrushin (1970), Gassando et al. (1978)
14) The theory of dynamical systems – Ornstein (1971), (1985), Gray (1988)
15) Matrix theory – Olkin and Pukelheim (1982)
16) Linear programming—Hoffman (1961), Barnes and Hoffman (1985), Derigs, Goecke and Schrader (1986), Goldfarb (1985), Kleinschmidt, Lee and Schannath (1987)

References

17) Statistics—concordance, associativeness and correlation of random quantities—Kruskal (1958), Lehman (1966), Robbins (1975), Lai and Robbins (1976), Mallows (1972), Rinott (1973), Bartfai (1970), Whitt (1976), Cambanis et al. (1976), Cambanis and Simons (1982), Wolff and Schweizer (1981), Ruschendorf (1980, 1981, 1985a,b, 1989), Gaffke and Ruschendorf (1981, 1984), Cuesta and Matran (1989)
18) Marginal problems and doubly stochastic matrices—Beasley and Gibson (1975), Berkes and Phillip (1979), Brown (1965), Brown and Shiflett (1970), Brunaldi and Czima (1975), Deland and Shiflett (1980), Isbell (1955,1962), Douglas (1964, 1966), Dowson and Landau (1982), Kellerer (1961-88), Hoffmann-Jorgensen (1977), Major (1978), Mukerjee (1985), Shortt (1983-87), Vorob'ev (1962, 1963), Sudakov (1972-76), Levin (1977-86), Levin and Milyutin (1979), Haneveld (1985)
19) A Probability Metrics Approach for Reducing the Bias of Optimality Gap Estimators in Stochastic Programming – R. Stockbridge (2010)
20) Convergence in Distribution of Random Metric Measure Spaces – A. Greven (2009)

References

21) Hilbert Space Embeddings and Metrics on Probability Measures – B. K. Sriperumbudur (2010)
22) Ideals of Operators and the Metric Approximation Property – V. Lima (2004)
23) Non-parametric Estimation of Integral Probability Metrics – B. K. Sriperumbudur (2009)
24) Probability Metrics and their Applications – H. Mostafaei (2011)
25) Quantization for Probability Measures in the Prokhorov Metric - S. GRAF (2009)
26) Strong Ideal Convergence in Probabilistic Metric Spaces – C. Sencimen (2009)
27) Numerical Approximations to the Mass Transfer Problem on Compact Spaces – J. R. Gabriel (2010)
28) On Solutions to the Mass Transfer Problem – J. Gonzalez-Hernandez (2006)
29) On the Consistency of the Mass Transfer Problem – J. Gonzalez-Hernandez (2006)
30) On Duality for a Generalized Monge–Kantorovich Problem – Y. Olubummo (2004)
31) On Minimum Kantorovich Distance Estimators – F. Bassetti (2006)
32) On the Equality between Monge's Infimum and Kantorovich's Minimum in Optimal Mass Transportation – A. Pratelli (2007)

References

33) On the Kantorovich–Rubinstein Theorem – D.A. Edwards (2011)
34) Optimal and Better Transport Plans – M. Beiglböck (2009)
35) Optimal Transport Maps for Monge–Kantorovich Problem on Loop Groups – S. Fang (2007)
36) A Monge–Kantorovich Mass Transport Problem for a Discrete Distance – N. Igbida (2011)
37) Asymptotic Analysis of a Class of Optimal Location Problems – G. Bouchitté (2011)
38) Asymptotic Analysis of the Optimal Cost in Some Transportation Problems with Random Locations – G. L. Torrisi (2011)
39) From the Schrödinger Problem to the Monge–Kantorovich Problem – C. Léonard (2012)
40) Kantorovich Metric: Initial History and Little-Known Applications – A. M. Vershik (2006)
41) On a System of Partial Differential Equations of Monge–Kantorovich Type – G. Crasta (2007)

THANK YOU !