THEORY OF PROBABILITY AND ITS APPLICATIONS

Volume XXIX Number 4

THE MONGE-KANTOROVICH MASS TRANSFERENCE PROBLEM AND ITS STOCHASTIC APPLICATIONS

S. T. RACHEV

(Translated by Bernard Seckler)

Contents

1. Introduction

The Monge-Kantorovich problem (MKP) has been the center of attention of many specialists in various areas of mathematics for a long time—differential geometry (see [7], [50] and the references there); functional analysis [25], [34]; infinite-dimensional linear programming [6], [46], [40], [100]; probability theory [82], [84], [79], [20]; mathematical statistics [91], [141], [54]; information theory and cybernetics [5], [94], [96]; statistical physics [10], [69]; the theory of dynamical systems [145]; and matrix theory [52], [109], [113]. Currently, it is now appropriate to talk about the MKP as being a whole range of problems with applications to many mathematical theories that seem different at first glance. Entire schools have been formed developing different offshoots of the MKP by making use of diversified mathematical language.

It is very difficult to encompass thoroughly all of the problems and results in the MKP, which apparently forms an epoch in the development of a considerable range of mathematics. This review article discusses the contemporary state of the MKP mainly paying attention to the probability aspects of the problem.

2. Statement of the Monge–Kantorovich Problem

This section should be viewed as an introduction to the MKP and its related problems. There are five known versions of the MKP.

2.1. Monge optimization problem. In 1781, Monge formulated the following problem (see [112], [59], [60], [87], [24], [50]) 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 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?

For the significance of this problem in the development of differential geometry, see [7].

2.2. Kantorovich's mass transference problem. The abstract form of Kantorovich's problem is as follows.

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 $P_2(\cdot) = P_2(U \times \cdot)$. Evaluate the functional

$$\mathcal{A}_c(P_1, P_2) = \inf\left\{\int_{U\times U} c(x, y)P(dx, dy) \colon P \in \mathcal{P}(P_1, P_2)\right\}, \tag{2.1}$$

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

We shall call the functional (2.1) *Kantorovich's functional.*

The measures P_1 and P_2 may be viewed as the initial and final distribution of mass and $\mathcal{P}(P_1, P_2)$ as the space of admissible transference plans. If the infimum in (2.1) is realized for some measure $P^* \in \mathcal{P}(P_1, P_2)$, 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.

Problem 2.2 was first formulated and studied by Kantorovich for a compact U and $c = d$ (see [23], [24]). It was shown that

$$\mathcal{A}_d(P_1, P_2) = \mathcal{B}_d(P_1, P_2), \tag{2.2}$$

where $\mathcal{B}_d$ is the Kantorovich metric in the space $\mathcal{P}_U$ of Borel probability measures on (U, d), namely,

$$\begin{gathered}\mathcal{B}_d(P_1, P_2) = \sup\left\{\left|\int_U f d(P_1 - P_2)\right| \colon f \in \mathrm{Lip}_{1,1}(U)\right\},\\ \mathrm{Lip}_{1,\alpha}(U) = \{f\colon U \to R^1\colon |f(x) - f(y)| \leqq \alpha d(x, y),\ x, y \in U,\ \sup_{x\in U} |f(x)| < \infty\}.\end{gathered} \tag{2.3}$$

Problem (2.2) with a continuous cost function on any compact space U was studied by Levin [30]-[33] and Levin and Milyutin [34] (see also [6], [46]).

Kantorovich's formulation differs from the Monge problem in that the class $\mathcal{P}(P_1, P_2)$ is broader than the class of one-to-one trnasference plans in Monge's sense (see [6], [50]). Sudakov [50] showed that if the 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.

In the case of a Polish (complete, separable, metric) space, the dual relation (2.2) is proved in the papers of Szulga [134], [135], Fernique [80], Hubert ([54] where d is a bounded metric), and Dudley-De Acosta [57]. Kantorovich's theorem for an arbitrary s.m.s. U may be stated as follows (see [43], [44]).

Let $\mathfrak{C}_1$ be the class of all functions $c(x, y) = H(d(x, y))$, $x, y \in U$, where the function H belongs to the class $\mathcal{H}_1$ of all nondecreasing continuous functions on $[0, \infty)$ for which $H(0) = 0$ and which satisfy Orlicz' condition

$$c_H = \sup\{H(2t)/H(t)\colon t > 0\} < \infty.$$

We also write $\mathcal{H}_2$ for the subset of all convex functions in $\mathcal{H}_1$ and $\mathfrak{C}_2$ for the set $\{H \circ d\colon H \in \mathcal{H}_2\}$.

Theorem 1. *Let* $c \in \mathfrak{C}_2$ *and* $\mathcal{A}_c(P_1, P_2)$ *be given by* (2.1). *Let*

$$\operatorname{Lip}^c(U) = \left\{(f, g) \in \bigcup_{\alpha>0} [\operatorname{Lip}_{1,\alpha}(U)]^{\times 2}\colon f(x) + g(y) \leqq c(x, y),\ x, y \in U\right\}$$

and

$$\mathcal{B}_c(P_1, P_2) = \sup\left\{\int_U f\,dP_1 + \int_U g\,dP_2\colon (f, g) \in \operatorname{Lip}^c(U)\right\}.$$

(i) *If* $\int_U c(x, a)(P_1 + P_2)(dx) < \infty$ *for some* $a \in U$, *then*

(2.4) $$\mathcal{A}_c(P_1, P_2) = \mathcal{B}_c(P_1, P_2).$$

(ii) *If* $\int_U d(x, a)(P_1 + P_2)(dx) < \infty$ *for some* $a \in U$, *then*

(2.5) $$\mathcal{A}_d(P_1, P_2) = \mathcal{B}_d(P_1, P_2),$$

where $\mathcal{B}_d$ *is given by formula* (2.3).

If P_1 *and* P_2 *are tight measures, then the infimum in* (2.1) *is attained.*

Recall that a measure P is said to be *tight* if for any positive ε there is a compact set K_ε such that $P(K_\varepsilon) > 1 - \varepsilon$ (see, for example, [79]).

The proof of Theorem 1 is based on the idea in the proof of the more general Theorem 3 given below.

The theorem implies that if $\mathfrak{A}$ is a class of pairs (f, g) of measurable functions satisfying $f(x) + g(y) \leqq c(x, y)$ for all $x, y \in U$ and $\mathfrak{A} \supset \operatorname{Lip}^c(U)$, then

$$\mathcal{A}_c(P_1, P_2) = \sup\left\{\int_U f\,dP_1 + \int_U g\,dP_2\colon (f, g) \in \mathfrak{A}\right\}$$

(see also [143] for U a Polish space and [100] for (U, d) an arbitrary space).

Dobrushin [10] proved that if $c(x, y) = I\{x \neq y\}$ is the indicator metric, then in a metric (not necessarily separable) space (U, d)

(2.6) $$\mathcal{A}_c(P_1, P_2) = \sup\{|P_1(A) - P_2(A)|\colon A \in \mathfrak{B}(U)\},$$

where $\mathfrak{B}(U)$ is a σ-algebra of Borel sets.

Relations (2.2), (2.4)-(2.6) furnish a complete description of the minimal metrics

$$(2.7) \qquad \mathbf{1}_p(P_1, P_2) = \inf\{\mathscr{L}_p(P) : P \in \mathscr{P}(P_1, P_2)\}, \qquad 0 \leqq p \leqq \infty,$$

with respect to functionals called compound probability metrics (see [14], 20])

$$\mathscr{L}_p(P) = \left[\int_{u\times U} d^p(x, y) P(dx, dy)\right]^{p'}, \qquad 0 < p < \infty, p' = \min(1, 1/p),$$

$$\mathscr{L}_0(P) = \int_{U\times U} I\{x \neq y\} P(dx, dy),$$

$$\mathscr{L}_\infty(P) = \operatorname{ess\,sup}_P d(x, y) \stackrel{\text{def}}{=} \inf\{\varepsilon > 0 : P(d(x, y) > \varepsilon) = 0\}.$$

It will be recalled that a s.m.s. (U, d) is called universally measurable (u.m.) if any measure $P \in \mathscr{P}(U)$ is tight. If (U, d) is u.m., then the functionals $\mathbf{1}_p$, $0 \leqq p \leqq \infty$, are metrics by virtue of the Fréchet-Strassen theorem on the existence of a measure with given marginal distributions (see [83], [63], [72], [53], [133], [48]-[50], [57], [132], [93], [153]). The problem of finding $\mathbf{1}_p$, $p > 1$, is known as Dudley's problem [79]. The dual relations for (2.7) when $0 \leqq p < \infty$ follow from the representations (2.2), (2.4)-(2.6); the relation

$$\mathbf{1}_\infty(P_1, P_2) = \inf\{\varepsilon > 0 : P_1(A) \leqq P_2(A^\varepsilon), A \in \mathfrak{B}(U)\},$$
$$A^\varepsilon = \{x : d(x, A) < \varepsilon\}$$

(see [77], [79], [38], [21]) is a simple consequence of Strassen's theorem

$$(2.8) \qquad \inf\{\mathscr{K}\mathscr{F}(P) : P \in \mathscr{P}(P_1, P_2)\} = \pi(P_1, P_2),$$

where $\mathscr{K}\mathscr{F}(P) = \inf\{\varepsilon > 0 : P(d(x, y) > \varepsilon) < \varepsilon\}$ is the Ky Fan (Fan' Tsi) distance and $\pi(P_1, P_2)$ is the Lévy-Prokhorov distance (see [13], [36], [133], [77], [79], [86], [135]). Dobrushin's theorem (2.6) also follows from Strassen's theorem (see [37]). In addition, (2.6) is a direct consequence of relation (2.5) of Theorem 1 in a s.m.s. (U, d^p), $0 \leqq p \leqq 1$, if $p \to 0$ (see [14]).

We now consider the topological structure of the functionals (2.1). Strassen's theorem (2.8) shows that a metric which is minimal with respect to distance in probability induces weak convergence. Hence, it is natural to expect that the metrics $\mathbf{1}_p$, $0 < p < \infty$, would also possess striking topological properties. Kantorovich and Rubinshtein [27] showed that $\mathbf{1}_1 = \mathscr{B}_d$ induces weak convergence when U is a compact space. Dudley [76], [79] (see also [54]) considers essentially the metric $\mathscr{B}_d$ of (2.3) with a bounded metric d and establishes the topological equivalence of $\mathscr{B}_d$ and π in a s.m.s. U. Dobrushin [10] proved that the weak convergence of the sequence $\{P_n, n = 1, 2, \cdots\} \subset \mathscr{P}_U$ to $P \in \mathscr{P}_U$ ($P_n \xrightarrow{w} P$) and the relation

$$\lim_{m\to\infty} \sup_n \int_U d(x, a) I\{d(x), a) > m\} P_n(dx- = 0$$

lead to $\lim_{n\to\infty} \mathbf{1}_1(P_n, P) = 0$ (see [57] for the special case of a Banach space U). The next assertion generalizes the results of Kantorovich-Rubinshtein, Dobrushin and Dudley on the topological structure of the functionals $\mathscr{A}_c$ in a s.m.s. U.

Theorem 2. *Let*

$$c \in \mathfrak{C}_1 \quad and \quad \int_U c(x, a) P_n(dx) < \infty, \qquad n = 0, 1, \cdots.$$

Then

$$\lim_{n\to\infty} \mathscr{A}_c(P_n, P_0) = 0 \Leftrightarrow P_n \xrightarrow{w} P_0, \lim_{n\to\infty} \int_U c(x, b)(P_n - P_0)(dx) = 0 \tag{2.9}$$

for some (and therefore for any) $b \in U$.

Theorem 2 is proved in [117], [43] and the special cases of it: $H(t) = t^p$, $p \geqq 1$, in [11], [116]; $p \geqq 1$, d a bounded metric, in [93]; and $p = 2$, $U = R^1$, $d(x, y) = |x - y|$ in [111]. The proof of (2.9) is based on a relation between minimal metrics (see [14]). Generalizations of (2.9) were considered in [43], [45]. For any $P_0 \in \mathscr{P}_U$ (U a Polish space), the space of measures $\{P: \mathscr{A}_c(P, P_0) < \infty\}$ is complete with respect to the metric functional $\mathscr{A}_c$, $c \in \mathfrak{C}_1$ (see [10], [41], [43]). In particular, the space $\{P: \mathbf{1}_p(P, P_0) < \infty\}$ is complete with respect to $\mathbf{1}_p$, $0 \leqq p \leqq \infty$ (see [93] for $1 \leqq p \leqq \infty$, $P_0(a\} = 1$).

In 1957, Kantorovich and Rubinshtein (see [26], [27]) studied the problem of transferring masses in the case the transit conveyances have been resolved, i.e., of determining the quantity

$$\mathscr{A}'_c(P_1, P_2) = \inf\left\{\int_{U\times U} c(x, y) Q(dx, dy) \colon Q \in \mathscr{P}'(P_1 P_2)\right\}, \tag{2.10}$$

where the space of admissible conveyances $\mathscr{P}'(P_1, P_2)$ consists of all bounded non-negative Borel measures on $U \times U$ satisfying

$$Q(A \times U) - Q(U \times A) = P_1(A) - P_2(A)$$

for all $A \in \mathfrak{B}(U)$. For U a compact space and $c(x, y)$ an arbitrary continuous cost function, Levin and Milyutin [34] proved the dual relation

$$\mathscr{A}'_c(P_1, P_2) = \mathscr{B}'_c(P_1, P_2), \tag{2.11}$$

where

$$\mathscr{B}'_c(P_1, P_2) \equiv \sup\left\{\int_U f\,d(P_1 - P_2) \colon f \colon U \to R^1, f(x) - f(y) \leqq c(x, y),\ x, y \in U\right\}.$$

The relation (2.11) continues to hold if U is a s.m.s., $c(x, y) = d(x, y) T(d(x, a), d(y, a))$, where a is some fixed point of U and $T(t, s) = T(s, t)$ is a continuous function on $(t \geqq 0, s \geqq 0)$ which is non-decreasing in this region in both its arguments and strictly positive everywhere with the possible exception of the lines $t = 0$ and $s = 0$ (see [43] and [79] (for $T \equiv 1$)). Clearly, $\mathscr{A}_c \geqq A'_c$ and if $c = d$, then $\mathscr{A}_c = \mathscr{A}'_c$ (see [27] for U compact and [57] for U Polish). Levin [31] proved that if (U, d) is compact, $c(x, x) = 0$, $c(x, y) \geqq 0$, and $c(x, y) + c(y, x) > 0$ for $x \neq y$, then $\mathscr{A}_c = \mathscr{A}'_c$ if and only if $c(x, y) + c(y, x) = d(x, y)$. In the following example, the differences between the functionals $\mathscr{A}_c$ and $\mathscr{A}'_c$ are especially noticeable.

EXAMPLE. Let $U=R^1$ and $c(x, y)=|x-y|h(\max(|x-a|, |y-a|)$, where $h(t)$ is a nondecreasing non-negative continuous function on the half-line $[0, \infty)$. Then

$$\mathscr{A}'_c(P_1, P_2)=\int_{-\infty}^{\infty}|F_1(x)-F_2(x)|h(|x-a|)\,dx,$$

$$\mathscr{A}_c(P_1, P_2)=\int_0^1 c(F_1^{-1}(t), F_2^{-1}(t))\,dt,$$

where F_j is the distribution function corresponding to P_j and F_j^{-1} is its inverse, $j=1, 2$.

The functional $\mathscr{A}'_c$ is frequently used in mathematical-economic models (see [46], [34], [144], [155]) but it is not applied in probability problems. Observe however the following relationship between the Fortet-Mourier metric and relation (2.11). In 1953, Fortet and Mourier [82] introduced a metric which is topologically equivalent to $\mathbf{l}_p$, $p\geqq 1$ (see [76]-[79], [116], [117], [43]) given by

$$\mathscr{FM}_p(P_1, P_2)=\sup\left\{\left|\int_U f\,d(P_1-P_2)\right|: f\in \mathfrak{G}^p\right\}$$

and

$$\mathfrak{G}^p\equiv\left\{g\colon U\to R^1, \sup_{r\geqq 1} r^{1-p}\sup\left\{\frac{|g(x)-g(y)|}{d(x, y)}: x\neq y, d(x, a)\leqq r, d(y, a)\leqq r\right\}\leqq 1\right\}$$

and a is a fixed point of the s.m.s. U. The functional $\mathscr{FM}_p$ has a dual representation of the form (2.11),

$$\mathscr{FM}_p(P_1, P_2)=\inf\left\{\int_{U\times U} D_p(x, y)P(dx, dy): P\in \mathscr{P}'(P_1, P_2)\right\},$$

$$D_p(x, y)\equiv d(x, y)\max[1, d^{p-1}(x, a), d^{p-1}(y, a)], \qquad x, y\in U.$$

2.3. Gini's measure of discrepancy. Already at the beginning of this century, the following question arose among probabilists: What is the proper way to measure the degree of difference between two random quantities (see the review article [101])? Specific contributions to the solution of this problem, which is closely related to Kantorovich's problem 2.2, were made by Gini (see [88]-[92]), Hoeffding [98], Fréchet [84] and by their successors (see [4], [35], [56], [61]-[68], [70]-[72], [81], [97], [102]-[111], [113], [115], [120]-[128], [131], [136], [138], [140]). In 1914, Gini [88], [89] introduced the concept of "simple measure of discrepancy" which coincides with Kantorovich's metric $\mathscr{A}_d(U=R^1, d(x, y)=|x-y|)$. Namely, Gini [88]-[90] studied the functional

$$\mathscr{K}(F_1, F_2)\equiv\inf\left\{\int_{R^2}|x-y|\,dF(x, y): F\in \mathscr{F}(F_1, F_2)\right\} \tag{2.12}$$

in the space $\mathscr{F}$ of one-dimensional distribution functions (d.f.) F_1 and F_2. In (2.12), $\mathscr{F}(F_1, F_2)$ is the class of all two-dimensional d.f. F with fixed marginal distributions $F_1(x)=F(x, \infty)$ and $F_2(x)=F(\infty, x)$, $x\in R^1$. Gini and his students devoted a great deal of study to the properties of the sample measure of discrepancy, Glivenko's theorem and goodness-of-fit tests in terms of $\mathcal{K}$ (see [61], [63]-[66], [70]-[72], [81], [88]-[92], [97], [103]-[106], [110], [115], [122]-[127]). Of especial importance in these investigations was the question of finding explicit expressions for the measure of discrepancy and its generalizations. Thus in 1943, Salvemini [123] showed that

$$\mathcal{K}(F_1, F_2) = \int_{-\infty}^{\infty} |F_1(x) - F_2(x)| \, dx \tag{2.13}$$

in the class of discrete d.f. and Dall'Aglio [70] extended (2.13) to all of $\mathcal{F}$. Formula (2.13) was proved and generalized in many ways (see for example, [27], [4], [22], [14], [142], [21], [79], [38], [118], [67]-[68], [136]). For example, by virtue of the Hoeffding-Fréchet inequalities (see [98], [83])

$$\begin{aligned} &\underline{F}(x, y) \leqq F(x, y) \leqq \bar{F}(x, y), \; F \in \mathcal{F}(F_1, F_2), \\ &\underline{F}(x, y) \equiv \max\{F_1(x) + F_2(y) - 1, 0\}, \\ &\bar{F}(x, y) \equiv \min\{F_1(x), F_2(y)\}, \qquad x, y \in R^1, \end{aligned}$$

any convex non-negative function φ on R^1 satisfies the relations

$$\begin{aligned} \inf\left\{\int_{R^2} \varphi \, dF \colon F \in \mathcal{F}(F_1, F_2)\right\} &= \int_{R^2} \varphi \, d\bar{F} = \int_0^1 \varphi(F_1^{-1}(u) - F_2^{-1}(u)) \, du, \\ \sup\left\{\int_{R^2} \varphi \, dF \colon F \in \mathcal{F}(F_1, F_2)\right\} &= \int_{R^2} \varphi \, d\underline{F} \\ &= \int_0^1 \varphi(F_1^{-1}(u) - F_2^{-1}(1-u)) \, du \end{aligned} \tag{2.14}$$

(see [68], [136]). In particular,

$$\mathbf{l}_p(P_1, P_2) = \left\{\int_0^1 |F_1^{-1}(t) - F_2^{-1}(t)|^p \, dt\right\}^{1/p}, \qquad p \geqq 1,$$

(see [98], [84], [107], [62] for $p=2$ and [63], [70], [72] for $p \geqq 1$) and

$$\mathbf{l}_\infty(P_1, P_2) = \sup\{|F_1^{-1}(x) - F_2^{-1}(x)| \colon x \in [0, 1]\}$$

(see [39]). The metric $\mathbf{l}_2$ (see [83], [84]) is sometimes called Fréchet distance (see [71]-[73]). Let $\mathcal{F}^*(F_1, F_2)$ denote the subset of $\mathcal{F}(F_1, F_2)$ of two-dimensional d.f. for which the infimum in (2.12) is attained. If $H \in \mathcal{F}(F_1, F_2)$ and $H(x, x) = \bar{F}(x, x)$, $x \in R^1$, then $H \in \mathcal{F}^*(F_1, F_2)$ (see [72], [22]). The following example shows that the class $\mathcal{F}^*(F_1, F_2)$ need not be a singleton set.

EXAMPLE. Let F_1 and F_2 be discrete d.f. for which there exist a positive ε and points $x_2 > x_1$ and $y_2 > y_1$ such that

$$F_1(x_i + 0) - F_1(x_i - 0) \geqq \varepsilon \quad \text{and} \quad F_2(y_i + 0) - F_2(y_i - 0) \geqq \varepsilon, \qquad i = 1, 2.$$

Then the measure $\bar{P}$ induced by $\bar{F}$ has a positive mass ε at the points (x_i, y_i),

$i=1,2$. Let P be a discrete measure on R^2 defined by the conditions:

$$P\{(x_i, y_i)\} = \bar{P}\{(x_i, y_i)\} - \varepsilon, \qquad i=1,2,$$
$$P\{(x_1, y_2)\} = \bar{P}\{(x_1, y_2)\} + \varepsilon,$$
$$P\{(x_2, y_1)\} = \bar{P}\{(x_2, y_2)\} + \varepsilon$$

and

$$P\{(x, y)\} = \bar{P}\{(x, y)\}$$

at all remaining points. Such a measure exists (cf. [136]) and its d.f. F belongs to $\mathscr{F}^*(F_1, F_2)$.

Other examples are discussed in Sudakov's paper [50].

The problem of describing the set of all optimal transference plans was studied in [23]-[24], [50], [74]-[75]. We point out that obtaining explicit expressions for the Kantorovich metric in the multi-dimensional case $U=R^n$, $n>1$, is an open question. We mention two results on the evaluation of $\mathbf{l}_2(P, Q)$, where P and Q are normal distributions in R^n with the means μ_P and μ_Q and real covariance matrices Σ_P and Σ_Q.

(i) If $\Sigma_P\Sigma_Q = \Sigma_Q\Sigma_P$, then (see [73])

$$\inf\left\{\int_{R^{2n}} \|\mathbf{x}-\mathbf{y}\|_2^2 \hat{P}(d\mathbf{x}, d\mathbf{y}) \colon \hat{P} \in \mathscr{P}(P, Q)\right\}$$
$$= \|\mu_P - \mu_Q\|_2^2 + \operatorname{tr}(\Sigma_P + \Sigma_Q - 2(\Sigma_P\Sigma_Q)^{1/2}),$$

where

$$\|\mathbf{x}-\mathbf{y}\|_2^2 = \sum_{i=1}^n |x_i - y_i|^2, \qquad \mathbf{x}=(x_1, \cdots, x_n), \qquad \mathbf{y}=(y_1, \cdots, y_n) \in R^n.$$

(The condition $\Sigma_P\Sigma_Q = \Sigma_Q\Sigma_P$ was not assumed to hold in [73] but as Shortt noted (in a private communication) without this condition the derivation of the formula is not legitimate.)

(ii) $$\inf\left\{\int_{R^{2n}} \|\mathbf{x}-\mathbf{y}\|_1^2 \hat{P}(d\mathbf{x}, d\mathbf{y}) \colon \hat{P} \in \mathscr{P}(P, Q)\right\}$$
$$= \|\mu_P - \mu_Q\|_1^2 + \operatorname{tr}(\Sigma_P + \Sigma_Q - 2(\sqrt{\Sigma_p}\Sigma_Q\sqrt{\Sigma_P})^{1/2},$$

where $\|\mathbf{x}-\mathbf{y}\|_1 = \sum_{i=1}^n |x_i - y_i|$ (see [93]).

The question of estimating the Kantorovich metric arises in many problems on the stability of stochastic models (see [14]-[20], [141]-[142]) but it is of especial interest in solving the problem of the uniqueness of a Gibbsian random field and in problems of phase transitions in statistical physics (see [8]-[10], [69], [5], [51], [55], [146]-[152], [154]), in particular, where $U=R^n$ and $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_1$.

In [11], [14], [16]-[20], [141]-[142], an estimate is given for the Kantorovich metric with the help of the first pseudomoment when U is a Banach space with norm $\|\cdot\|$:

$$\mathbf{l}_1(P, Q) = \inf\left\{\int\!\!\int_{U\times U} \|\mathbf{x}-\mathbf{y}\|\hat{P}(d\mathbf{x}, d\mathbf{y}) \colon \hat{P} \in \mathscr{P}(P, Q)\right\}$$

$$\leqq \nu_1(P, Q) = \int_U \|\mathbf{x}\|\,|P-Q|\,d\mathbf{x}.$$

The next three estimates refine the upper bound when $U = R^n$ and $\|\cdot\| = \|\cdot\|_1$. For $n = 1$, they become equalities.

(i) Let P and Q have the densities p and q. Then

$$\mathbf{l}_1(P, Q) \leqq \alpha_1(P, Q),$$

where

$$\alpha_1(P, Q) = \int_{R^n} \|\mathbf{x}\| \left|\int_0^1 (p-q)(x_1/t, \cdots, x_n/t)t^{-n-1}\,dt\right| dx_1 \cdots dx_n \leqq \nu_1(P, Q).$$

(ii) Let $(p_i - q_i)(x_1, \cdots, x_i) = \int_{R^{n-i}} (p-q)(x_1, \cdots, x_n)\,dx_{i+1} \cdots dx_n$. Then (see [42])

$$\mathbf{l}_1(P, Q) \leqq \alpha_2(P, Q),$$

where

$$\alpha_2(P, Q) = \int_{-\infty}^{\infty} \left|\int_{-\infty}^{t_1} (p_1 - q_1)(x_1)\,dx_1\right| dt_1$$

$$+ \sum_{i=2}^{n} \int_{R^{i-1}} \left[\int_{-\infty}^{0} \left|\int_{-\infty}^{t_i} (p_i - q_i)(x_1, \cdots, x_i)\,dx_i\right| dt_i\right.$$

$$\left. + \int_0^{\infty} \left|\int_{t_i}^{\infty} (p_i - q_i)(x_1, \cdots, x_i)\,dx_i\right| dt_i\right] dx_1 \cdots dx_{i-1} \leqq \nu_1(P, Q).$$

(iii) In the one-dimensional case, the Kantorovich metric

$$\mathscr{K}(F, G) = \inf\{\mathbf{E}_\Phi|X - Y| \colon \Phi \in \mathscr{F}(F, G)\} = \int_0^1 |F^{-1}(x) - G^{-1}(x)|\,dx$$

"is attained" on the pair of random variables $X^* = F^{-1}(V)$ and $Y^* = G^{-1}(V)$, where V is a uniformly distributed random variable in $(0, 1)$. In the multidimensional case, the Kantorovich metric has the form

$$\mathscr{K}(P, Q) = \inf\{\mathbf{E}_\Phi\|\mathbf{X}-\mathbf{Y}\| \colon \Phi \in \mathscr{F}^n(F, G)\},$$

where F and G are the respective d.f. of measures P and Q and $\mathscr{F}^n(F, G)$ is the space of all $2n$-dimensional d.f. Φ with fixed marginal distributions

$$\Phi(x_1, \cdots, x_n, \infty, \cdots, \infty) = F(x_1, \cdots, x_n),$$

$$\Phi(\infty, \cdots, \infty, x_1, \cdots, x_n) = G(x_1, \cdots\cdot, x_n)(x_1, \cdots, x_n \in R^1).$$

The following estimate is due to Kalashnikov. Let the random vectors $\mathbf{X} = (X_1, \cdots, X_n)$ and $\mathbf{Y} = (Y_1, \cdots, Y_n)$ have respective d.f. F and G. Put $\varphi(x_1) = \mathbf{P}(X_1 < x_1)$, $\varphi(x_{i+1};\, x_1, \cdots, x_i) = \mathbf{P}\,(X_{i+1} < x_{i+1} | X_1 = x_1, \cdots, X_i = x_i)$ and define functions ψ similarly for the vector $\mathbf{Y}$. Let $V_1, \cdots, V_n$ be independent r.v. distributed uniformly on $(0, 1)$. Let

$$X_1^* = \varphi^{-1}(V_1), \qquad X_2^* = \varphi^{-1}(V_2; X_1^*), \quad \cdots, \quad X_n^* = \varphi^{-1}(V_n; X_1^*, \cdots, X_{n-1}^*),$$
$$Y_1^* = \psi^{-1}(V_1), \qquad Y_2^* = \psi^{-1}(V_2; Y_1^*), \quad \cdots, \quad Y_n^* = \psi^{-1}(V_n; Y_1^*, \cdots, Y_{n-1}^*).$$

Then the random vectors $\mathbf{X}^* = (X_1^*, \cdots, X_n^*)$ and $\mathbf{Y}^* = (Y_1^*, \cdots, Y_n^*)$ have d.f. F and G and hence

$$\mathcal{K}(P, Q) \leqq \mathbf{E}\|\mathbf{X}^* - \mathbf{Y}^*\|.$$

In the case $\|\cdot\| = \|\cdot\|_1$ and $U = R^n$, the following lower bound exists for $\mathcal{K}(P, Q)$. Let the measure P have one-dimensional marginal d.f. $F_1, \cdots, F_n$ and Q the marginal d.f. $G_1, \cdots, G_n$. Then $\mathcal{K}(P, Q) \geqq \sum_{i=1}^n \int_{-\infty}^{\infty} |F_i(x) - G_i(x)|\, dx$, and equality is attained when P is uniquely determined by the collection $F_1, \cdots, F_n$ and Q by the collection $G_1, \cdots, G_n$ (see [136], [140]).

2.4. Ornstein distance. Let (U, d) be a s.m.s. and let $d_{n,\alpha}$, $\alpha \in [0, \infty]$, be the analogue of the Hamming metric $U^{\times n}$, namely,

$$d_{n,\alpha}(\mathbf{x}, \mathbf{y}) = \frac{1}{n}\left[\sum_{i=1}^{n} d^\alpha(x_i, y_i)\right]^{\alpha'}, \qquad \mathbf{x} = (x_1, \cdots, x_n) \in U^{\times n},$$

$\mathbf{y} = (y_1, \cdots, y_n) \in U^{\times n}, 0 < \alpha < \infty, \alpha' = \min(1, 1/\alpha),$

$$d_{n,0}(\mathbf{x}, \mathbf{y}) = \frac{1}{n}\sum_{i=1}^{n} I\{x_i \neq y_i\},$$

$$d_{n,\infty}(\mathbf{x}, \mathbf{y}) = \frac{1}{n}\max\{d(x_i, y_i) \colon i = 1, \cdots, n\}.$$

For any Borel probability measures P and Q on $U^{\times n}$, define the following analogue of the Kantorovich distance:

$$D_{n,\alpha}(P, Q) = \inf\left\{\int_{U^{\times 2n}} d_{n,\alpha}\, d\hat{P} \colon \hat{P} \in \mathcal{P}(P, Q)\right\}.$$

The distance $D_{n,0}$ is known among specialists in the theory of dynamical systems and coding theory as Ornstein's $\bar{d}$-distance (see [114], [94]-[96], [145]). In [94] (see also [136]), a generalization of the Ornstein distance is considered called the $\bar{\rho}$-distance which coincides with $D_{n,1}$. In information theory, the Kantorovich metric $D_{1,1}$ is known as the Vassershtein (sometimes Lévy-Vassershtein) distance (see [5], [10], [136]). It is possible to show that

$$\text{(2.15)} \quad \begin{aligned} D_{n,\alpha}(P, Q) &= \sup\left\{\left|\int f\, d(P - Q)\right| \colon f \colon U^{\times N} \to R^1, L_{n,\alpha}(f) \leqq 1\right\}, \\ L_{n,\alpha}(f) &= \sup\{|f(\mathbf{x}) - f(\mathbf{y})| / d_{n,\alpha}(\mathbf{x}, \mathbf{y}), \qquad \mathbf{x} \neq \mathbf{y}, \mathbf{y} \in U^{\times n}\} \end{aligned}$$

for all $\alpha \in [0, \infty)$ (see Theorem 3 below).

2.5. Multi-dimensional Kantorovich problem. Consider the following problem which generalizes all of the preceding statements.

Let $\tilde{P} = \{P_i, i = 1, \cdots, N\}$ be the set of probability measures given on a s.m.s. (U, d) and let $\mathfrak{P}(\tilde{P})$ be the space of all Borel probability measures P on the direct product $U^{\times N}$ with fixed projections P_i on the ith coordinates, $i = 1, \cdots, N$. Evaluate the functional

$$\mathbf{A}_c(\tilde{P}) = \inf\left\{\int_{U^{\times N}} c\, dP: P \in \mathfrak{P}(\tilde{P})\right\}, \tag{2.16}$$

where c is a given continuous function on $U^{\times N}$ (*see* [40], [109], [128], [136]).

This transport problem of infinite-dimensional linear programming is of interest in its own right in problems of stability of stochastic models (see [11], [141], [22], [40]). This is related to the fact that if $\{P_1^{(i)}, \cdots, P_N^{(i)}\}$, $i = 1, 2$, are two sets of probability measures on (U, d) and $P^{(i)} = P_1^{(i)} \times \cdots \times P_N^{(i)}$ are their products, then the value of the Kantorovich functional

$$\mathscr{A}_{c^*}(P^{(1)}, P^{(2)}) = \inf\left\{\int_{U^{\times 2N}} c^*\, d\hat{P}: \hat{P} \in \mathscr{P}(P^{(1)}, P^{(2)})\right\},$$

$$c^*(x_1, \cdots, x_N, y_1, \cdots, y_N) = \varphi(c_1(x_1, y_1), \cdots, c_N(x_N, y_N)),$$

$$x_i, y_i \in U, \quad i = 1, \cdots, N,$$

where φ is some continuous function on R^N, coincides with

$$\begin{aligned}\mathbf{A}_{c^*}&(P_1^{(1)}, \cdots, P_N^{(1)}, P_1^{(2)}, \cdots, P_N^{(2)})\\ &= \inf\left\{\int_{U^{\times 2N}} c^*\, dP: P \in \mathfrak{P}(P_1^{(1)}, \cdots, P_N^{(1)}, P_1^{(2)}, \cdots, P_N^{(2)}\right\}.\end{aligned}$$

When $U = R^1$, the following theorem of Lorentz (see [109], [128], [136], [40]) establishes an explicit representation for the functional $\mathbf{A}_c(\tilde{P})$. A function $W: R^2 \to R^1$ is said to be *2-antitone* if

$$W(x, y) + W(x', y) - W(x, y') - W(x', y) \leqq 0 \quad \text{for all } x \leqq x', y \leqq y'.$$

A function $W: R^N \to R^1$, $N \geqq 2$, is said to be *N-antitone* if W is 2-antitone in any two of its arguments (for different examples of N-antitone functions, see [68], [136], [40]). Let the probability measure P_i have a d.f. F_i, $i = 1, \cdots, N$, and let $\bar{H}(x) = \min\{F_i(x): i = 1, \cdots, N\}$. Then any continuous N-antitone function W satisfies the relation (see [136])

$$\int_{R^N} W\, d\bar{H} = \mathbf{A}_W(\tilde{P})$$

if and only if at least one of the following two conditions holds:

(a) $W \geqq h$ for some continuous function h satisfying

$$\left|\int_{R^N} h\, dP\right| < \infty, \qquad P \in \mathfrak{P}(\tilde{P});$$

(b) $\sup\{|\int_{R^N} W\, dP|: P \in \mathfrak{P}(\tilde{P})\} < \infty$.

3. Multi-Dimensional Kantorovich Theorem

In this section we shall prove the duality theorem for the multi-dimensional Kantorovich problem and we shall study the topological structure of the functional (2.16).

For brevity, $\mathcal{P}$ will denote the space $\mathcal{P}_U$ of all Borel probability measures on a s.m.s. (U, d). Let $N (N \geqq 2)$ be an integer and let $\|\mathbf{b}\|$ ($\mathbf{b} \in R^m$, $m = \binom{N}{2}$) be a monotone seminorm $\|\cdot\|$, i.e., $\|\cdot\|$ is a seminorm in R^m with the following property: if $0 < b_i' < b_i''$, $i = 1, \cdots, m$, then $\|\mathbf{b}'\| \leqq \|\mathbf{b}''\|$. For example, $\|\mathbf{b}\| = [\sum_{i=1}^m |b_i|^p]^{1/p}$, $\|\mathbf{b}\| = \max\{|b_i|: i = 1, \cdots, m\}$, $\|b\| = |\sum_{i=1}^m b_i|$ and $\|\mathbf{b}\| = [|\sum_{i=1}^k b_i|^p + |\sum_{i=k+1}^m b_i|^p]^{1/p}$, $p \geqq 1$. For any $\mathbf{x} = (x_1, \cdots, x_N) \in U^{\times n}$, let

$$\mathcal{D}(\mathbf{x}) = \|d(x_1, x_2), d(x_1, x_3), \cdots, d(x_1, x_N), d(x_2, x_3), \cdots, d(x_{N-1}, x_N)\|.$$

Let $\tilde{P} = (P_1, \cdots, P_N)$ be a finite set of measures in $\mathcal{P}$ and let

$$\mathbf{A}_D(\tilde{P}) = \inf\left\{\int_{U^{\times N}} D\, dP: P \in \mathfrak{P}(\tilde{P})\right\}, \tag{3.1}$$

where $D(\mathbf{x}) = H(\mathcal{D}(\mathbf{x}))$, $\mathbf{x} \in U^{\times N}$ and $H \in \mathcal{H}_2$.

Let $\mathcal{P}^H$ be the space of all measures in $\mathcal{P}$ for which $\int_U H(d(x, a))P(dx) < \infty$, $a \in U$. For any $U_0 \subseteq U$ define the class $\mathrm{Lip}(U_0) = \bigcup_{\alpha > 0} \mathrm{Lip}_{1,\alpha}(U_0)$, where $\mathrm{Lip}_{1,\alpha}(U_0) = \{f: U \to R^1: |f(x) - f(y)| \leqq \alpha\, d(x, y)$ for all $x, y \in U_0$, $\sup\{|f(x)|: x \in U_0\} < \infty\}$.

Define the class $\mathfrak{G}(U_0) = \{\mathbf{f} = (f_1, \cdots, f_N): \sum_{i=1}^N f_i(x_i) \leqq D\ (x_1, \cdots, x_N)$ for $x_i \in U_0, f_i \in \mathrm{Lip}(U_0), i = 1, \cdots, N\}$ and for any class $\mathfrak{A}$ of vectors $\mathbf{f} = (f_1, \cdots, f_N)$ of measurable functions, let

$$\mathbf{K}(\tilde{P}, \mathfrak{A}) = \sup\left\{\sum_{i=1}^N \int_U f_i\, dP_i: \mathbf{f} \in \mathfrak{A}\right\}$$

and hence the following inequality holds:

$$\mathbf{A}_D(\tilde{P}) \geqq \mathbf{K}(\tilde{P}; \mathfrak{G}(U)). \tag{3.2}$$

The next theorem (an extension of Kantorovich's theorem to the multi-dimensional case) shows that exact equality holds in (3.2).

Theorem 3. *For any s.m.s. (U, d) and for any set $\tilde{P} = (P_1, \cdots, P_n)$, $P_i \in \mathcal{P}^H$, $i = 1, \cdots, N$,*

$$\mathbf{A}_D(\tilde{P}) = \mathbf{K}(\tilde{P}; \mathfrak{G}(U)). \tag{3.3}$$

If the set $\tilde{P}$ consists of tight measures, then the infimum is attained in (3.1).

PROOF. I. Suppose first that d is a bounded metric in U and let

$$\rho_i(x_i, y_i) = \sup\{|D(x_1, \cdots, x_N) - D(y_1, \cdots, y_N)|: x_j = y_j \in U,\ j = 1, \cdots, N, j \neq i\} \tag{3.4}$$

for $x_i, y_i \in U$, $i = 1, \cdots, N$. Since H is a convex function, $\rho_1, \cdots, \rho_N$ are bounded metrics. Let $U_0 \subseteq U$ and let $\mathfrak{G}(U_0)$ be the space of all collections $\mathbf{f} = (f_1, \cdots, f_N)$ of measurable functions on U_0 such that $f_1(x_1) + \cdots + f_N(x_N) \leqq D(x_1, \cdots, x_N)$,

$x_1, \cdots, x_n \in U_0$. Let $\mathfrak{G}''(U_0)$ be a subset of $\mathfrak{G}'(U_0)$ of collections $\mathbf{f}$ for which $|f_i(x)-f_i(y)| \leqq \rho_i(x, y)$, $x, y \in U_0$, $i=1, \cdots, N$. We wish to show that if $P_i(U_0)=1$, $i=1, \cdots, N$, then

$$\mathbf{K}(\tilde{P}; \mathfrak{G}'(U_0)) = \mathbf{K}(\tilde{P}; \mathfrak{G}''(U)). \tag{3.5}$$

Let $\mathbf{f} \in \mathfrak{G}'(U_0)$. We define sequentially the functions

$$\begin{aligned} f_1^*(x_1) &= \inf\{D(x_1, \cdots, x_N) - f_2(x_2) - \cdots - f_N(x_N): \\ &\qquad x_2, \cdots, x_N \in U_0\}, x_1 \in U, \\ f_2^*(x_2) &= \inf\{D(x_1, \cdots, x_N) - f_1^*(x_1) - f_3(x_3) - \cdots - f_N(x_N): \\ &\qquad x_1 \in U, x_3, \cdots, x_N \in U_0\}, x_2 \in U, \cdots, \\ f_N^*(x_N) &= \inf\{D(x_1, \cdots, x_N) - f_1^*(x_1) - \cdots - f_{N-1}^*(x_{N-1}): \\ &\qquad x_1, \cdots, x_{N-1} \in U\}, x_N \in U. \end{aligned}$$

The collection $\mathbf{f}^* = (f_1^*, \cdots, f_N^*)$ belongs to the set $\mathfrak{G}''(U)$ and $f_i^* \geqq f_i(x)$ for all $i=1, \cdots, N$ and $x \in U_0$. Hence, $\sum_{i=1}^n \int_U f_i \, dP_i \leqq \sum_{i=1}^n \int_U f_i^* \, dP_i$ which implies the inequality

$$\mathbf{K}(\tilde{P}; \mathfrak{G}'(U_0)) \leqq \mathbf{K}(\tilde{P}; \mathfrak{G}''(U_0)) \tag{3.6}$$

from which (3.5) clearly follows.

CASE 1. Let U be a finite space with the elements $u_1, \cdots, u_n$. From the duality principle in linear programming, we have

$$\begin{aligned} \mathbf{A}_D(\tilde{P}) &= \inf\left\{ \sum_{i_1=1}^n \cdots \sum_{i_N=1}^n D(u_{i_1}, \cdots, u_{i_N}) \pi(i_1, \cdots, i_N): \right. \\ &\qquad \left. \pi(i_1, \cdots, i_N) \geqq 0, \sum_{i_j: j \neq k} \pi(i_1, \cdots, i_N) = P_k(u_{i_k}), k=1, \cdots, N \right\} \\ &= \sup\left\{ \sum_{i=1}^n \sum_{j=1}^N f_j(u_i) P_j(u_i): \sum_{j=1}^N f_j(u_j) \leqq D(u_1, \cdots, u_N), u_1, \cdots, u_N \in U \right\} \\ &= \mathbf{K}(\tilde{P}; \mathfrak{G}'(U)). \end{aligned}$$

Therefore (3.6) implies the chain of inequalities

$$\mathbf{K}(\tilde{P}; \mathfrak{G}(U)) \geqq \mathbf{K}(\tilde{P}; \mathfrak{G}''(U)) \geqq \mathbf{K}(\tilde{P}; \mathfrak{G}'(U)) \geqq \mathbf{A}_D(\tilde{P}),$$

from which (3.3) follows by virtue of (3.2).

CASE 2. Let U be a compact set. For any $n=1, 2, \cdots$, choose disjoint non-empty Borel sets $A_1, \cdots, A_{m_n}$ of diameter less than $1/n$ whose union is U. Define a mapping $h_n: U \to U_n = \{u_1, \cdots, u_{m_n}\}$ such that $h_n(A_i) = u_i$, $i=1, \cdots, m_n$. According to (3.5), we have for the collection $\tilde{P}_n = (P_1 \circ h_n^{-1}, \cdots, P_N \circ h_n^{-1})$ the relation

$$\mathbf{K}(\tilde{P}_n; \mathfrak{G}'(U_n) = \sup\left\{ \sum_{i=1}^n \int_U f_i(h_n(u)) P_i(du): \mathbf{f} \in \mathfrak{G}'(U) \right\}. \tag{3.7}$$

If $\mathbf{f}\in\mathfrak{G}'(U)$, then $\sum_{i=1}^n f_i(h_n(u_i))\leqq D(u_1,\cdots,u_n)+K/n$, where the constant K is independent of n and $u_1,\cdots,u_n\in U$. Hence, from (3.7) we have

$$\mathbf{K}(\tilde{P}_n;\mathfrak{G}'(U_n))\leqq\mathbf{K}(\tilde{P};\mathfrak{G}'(U))+K/n. \tag{3.8}$$

According to Case 1, there exists a measure $P^{(n)}\in\mathfrak{P}(\tilde{P}_n)$ such that

$$\int_{U^{\times N}} D\,dP^{(n)}=\mathbf{K}(\tilde{P}_n;\mathfrak{G}'(U_n)). \tag{3.9}$$

Since $P_i\circ h_n^{-1}$ converges weakly to P_i, $i=1,\cdots,N$, the sequence $\{P^{(n)}, n=1,2,\cdots\}$ is weakly compact. Let P^* be a limit of it in the sense of weak convergence. From (3.8) and (3.9) it follows that

$$\int_{U^{\times N}} D\,dP^*\leqq\mathbf{K}(\tilde{P};\mathfrak{G}'(U)).$$

CASE 3. Let (U,d) be a bounded s.m.s. Since

$$\int_U H(d(x,a))P_i(dx)<\infty,$$

the convexity of H and (3.4) imply that

$$\int_U \rho_i(x,a)P_i(dx)<\infty,\qquad i=1,\cdots,N.$$

Let the P_i be tight measures. Then for each $n=1,2,\cdots$ there exists a compact set K_n such that

$$\sup_{1\leqq i\leqq N}\int_{U\setminus K_n}(1+\rho_i(x,a))P_i(dx)<1/n. \tag{3.10}$$

For any $A\in\mathfrak{B}(U)$, put

$$P_{i,n}(A)=P_i(A\cap K_n)+P_i(U\setminus K_n)\delta_a(A),\qquad \tilde{P}_n=(P_{1,n},\cdots,P_{N,n}),$$

where

$$\delta_a(A)=\begin{cases}1, & a\in A,\\ 0, & a\notin A,\end{cases}$$

is the indicator function of the set A. By (3.5),

$$\begin{aligned}&\mathbf{K}(\tilde{P}_n;\mathfrak{G}'(K_n\cup\{a\}))\\ &\leqq\sup\left\{\sum_{i=1}^N\int_U f_i(x)P_i(dx)+\int_{U\setminus K_n}\rho_i(x,a)P_i(dx)\colon \mathbf{f}\in\mathfrak{G}(U)\right\}\\ &\leqq\mathbf{K}(\tilde{P};\mathfrak{G}(U))+N/n.\end{aligned} \tag{3.11}$$

According to Case 2, there exists a measure $P^{(n)}\in\mathfrak{P}(\tilde{P})$ such that

$$\int_{U^{\times N}} D\,dP^{(n)}\leqq\mathbf{K}(\tilde{P}_n;\mathfrak{G}'(K_n\cup\{a\}). \tag{3.12}$$

We then obtain (3.3) from relations (3.11) and (3.12) similarly to Case 2.

Now let $P_1, \cdots, P_N$ be measures that are not necessarily tight (see [79], [40]). Let $\bar{U}$ be the completion of U. To any positive ε, choose the largest set A such that $d(x, y) \geqq \varepsilon/2$ for all $x \neq y$, $x, y \in A$. The set A is countable: $A = \{x_1, x_2, \cdots\}$. Let $\bar{A}_n = \{x \in \bar{U}: d(x, x_n) < \varepsilon/2 \leqq d(x, x_j) \text{ for all } j < n\}$ and let $A_n = \bar{A}_n \cap U$. Then $\bar{A}_n$, $n = 1, 2, \cdots$, are disjoint Borel sets in $\bar{U}$ and A_n, $n = 1, 2, \cdots$, are disjoint sets in U of diameter less than ε. Let $\bar{P}_i$ be the measure generated on $\bar{U}$ by P_i, $i = 1, \cdots, N$. Then for $\mathbf{Q} = (\bar{P}_1, \cdots, \bar{P}_N)$ there exists a measure $\bar{\mu} \in \mathfrak{P}(\mathbf{Q})$ such that

$$\int_{\bar{U}^{\times N}} D\, d\bar{\mu} = \mathbf{K}(\mathbf{Q}; \mathfrak{G}(U)).$$

Let $P_{i,m}(B) = P_i(B \cap A_m)$ for all $B \in \mathfrak{B}(U)$, $i = 1, \cdots, N$. To any multiple index $\mathbf{m} = (m_1, \cdots, m_N)$, $m_i = 1, 2, \cdots$, $i = 1, \cdots, N$, define the measure (see [79], [40])

$$\mu_{\mathbf{m}} = c_{\mathbf{m}} P_{1,m_1} \times \cdots \times P_{N,m_N},$$

where the constant $c_{\mathbf{m}}$ is chosen so that $\mu_{\mathbf{m}}(A_{m_1} \times \cdots \times A_{m_N}) = \bar{\mu}(A_{m_1} \times \cdots \times A_{m_n})$. Let $\mu_\varepsilon = \sum_{\mathbf{m}} \mu_{\mathbf{m}}$. Then $\mu_\varepsilon \in \mathfrak{P}(\tilde{P})$ and, to each positive ε,

$$\begin{aligned}
&\mu_\varepsilon(\mathscr{D}(y_1, \cdots, y_N) > \alpha + 2\varepsilon\|\mathbf{e}\|) \\
&\quad \leqq \sum_{\mathbf{m}} \{\mu_{\mathbf{m}}(A_{m_1} \times \cdots \times A_{m_n}): \mathscr{D}(x_1, \cdots, x_N) > \alpha + \varepsilon\|\mathbf{e}\|\} \\
&\quad \leqq \bar{\mu}(\mathscr{D}(y_1, \cdots, y_N) > \alpha),
\end{aligned}$$

where $\mathbf{e}$ is a unit vector in R^m. Since $H(t)$ is strictly increasing and $D(x) = H(\mathscr{D}(x))$,

$$\begin{aligned}
\int_{U^{\times N}} D(\mathbf{x})\mu_\varepsilon(d\mathbf{x}) &= \int_0^\infty \mu_\varepsilon(\mathscr{D}(\mathbf{x}) > t)\, dH(t) \\
&\leqq \int_0^\infty \bar{\mu}(\mathscr{D}(\mathbf{x}) > t)\, dH(t + 2\varepsilon\|\mathbf{e}\|) + H(2\varepsilon\|\mathbf{e}\|) \\
&\leqq \int_{\bar{U}^{\times N}} D(\mathbf{x})\bar{\mu}(d\mathbf{x}) + \int_{\bar{U}^{\times N}} (H(\mathscr{D}(\mathbf{x}) + 2\varepsilon\|\mathbf{e}\|) - D(\mathbf{x}))\bar{\mu}(d\mathbf{x}) \\
&\quad + H(2\varepsilon\|\mathbf{e}\|).
\end{aligned}$$

From the Orlicz condition, it follows that for any positive p, the inequality

$$\begin{aligned}
&\int_{U^{\times N}} (H(D(\mathbf{x}) + 2\varepsilon\|\mathbf{e}\|) - D(\mathbf{x}))\bar{\mu}(d\mathbf{x}) \\
&\quad \leqq \sup\{H(t + 2\varepsilon\|\mathbf{e}\|) - H(t): t \in [0, 2p\|\mathbf{e}\|]\} \\
&\quad\quad + c_1 \sum_{i=1}^N \int_U H(d(x, a)) I\{d(x, a) > p/N\} P_i(dx)
\end{aligned}$$

holds, where c_1 is a constant independent of ε and p. As $\varepsilon \to 0$ and $p \to \infty$, we obtain

$$\limsup_{\varepsilon \to 0} \int_{U^{\times N}} D\, d\mu_\varepsilon \leqq \int_{\bar{U}^{\times N}} D\, d\bar{\mu} = \mathbf{K}(\mathbf{Q}; \mathfrak{G}(\bar{U})) = \mathbf{K}(\tilde{P}; \mathfrak{G}(U)).$$

II. Let U be any s.m.s. Suppose that $P_1, \cdots, P_N$ are tight measures. For any $n=1,2,\cdots$, define the bounded metric $d_n = nd/(n+d)$. Write $D_n(x_1,\cdots,x_N) = H(\|d_n(x_1,x_2),\cdots,d_n(x_1,x_N),d_n(x_2,x_3),\cdots,d_n(x_{N-1},x_N)\|)$. According to part I of the proof, there exists a measure $P^{(n)} \in \mathfrak{P}(\tilde{P})$ such that

$$\int_{U^{\times N}} D_n\, dP^{(n)} = \mathbf{K}(\tilde{P}; \mathfrak{G}(U, d_n)). \tag{3.13}$$

Since $P^{(n)}$, $n=1,2,\cdots$, is a uniformly tight sequence, on passing to a subsequence if necessary, we may assume that $P^{(n)}$ converges weakly to $P^{(0)} \in \mathfrak{P}(\tilde{P})$. By Skorokhod's theorem (see [47], [77], [79]), there exist a probability space (Ω, μ) and a sequence $\{X_k, k=0,1,\cdots\}$ of N-dimensional random vectors defined on (Ω,μ) and assuming values on $U^{\times N}$. Moreover, for any $k=0,1,\cdots$, the vector X_k has distribution $P^{(k)}$ and the sequence $X_1, X_2, \cdots$ converges μ-almost everywhere to X_0. According to (3.13), we have $\mathbf{K}(\tilde{P}; \mathfrak{G}(U, d_k)) = \int D_k(X_k)\, d\mu \to \int D(X_0)\, d\mu$. Hence

$$\mathbf{K}(\tilde{P}; \mathfrak{G}(U)) \geqq \lim_{k\to\infty} \mathbf{K}(\tilde{P}; \mathfrak{G}(U, d_k)) \geqq \mathbf{A}_D(\tilde{P}),$$

which by virtue of (3.2) implies (3.3). The theorem is proved.

As already mentioned, the multi-dimensional Kantorovich theorem can be interpreted naturally as a criterion for the closeness of n-dimensional sets of probability measures. Let (U_i, d_i) be a s.m.s., and $P_i, Q_i \in \mathcal{P}_{U_i}$, $i=1,\cdots,n$. Write $\tilde{P}=(P_1,\cdots,P_n)$, $\tilde{Q}=(Q_1,\cdots,Q_n)$, P_i, $Q_i \in \mathcal{P}_{U_i}$ and $\Delta(\mathbf{x},\mathbf{y}) = H(\|d_1(x_1,y_1),\cdots,d_n(x_n,y_n)\|)$, where $\mathbf{x}=(x_1,\cdots,x_n)$ and $\mathbf{y}(y_1,\cdots,y_n) \in U_1\times\cdots\times U_n = \mathfrak{A}$ and $\|\cdot\|_n$ is a monotone seminorm in R^n. The analogue of the Kantorovich distance in $\tilde{\mathcal{P}} = \mathcal{P}_{U_1}\times\cdots\times\mathcal{P}_{U_n}$ is defined as follows:

$$\mathfrak{K}_H(\tilde{P},\tilde{Q}) = \inf \int_{\mathfrak{A}\times\mathfrak{A}} \Delta(\mathbf{x},\mathbf{y}) P(d\mathbf{x}\, d\mathbf{y}) \colon P \in \mathfrak{P}(\tilde{P},\tilde{Q}), \tag{3.14}$$

where $\mathfrak{P}(\tilde{P},\tilde{Q})$ is the space of all probability measures on $\mathfrak{A}\times\mathfrak{A}$ with fixed one-dimensional marginal distributions $P_1,\cdots,P_n, Q_1,\cdots,Q_n$. The Orlicz condition implies that $\mathfrak{K}_H$ satisfies the following analogue of the triangle inequality:

$$\mathfrak{K}_H(\tilde{P}_1,\tilde{P}_2) \leqq c_H[\mathfrak{K}_H(\tilde{P}_1,\tilde{P}_3) + \mathfrak{K}_H(\tilde{P}_3,\tilde{P}_2)].$$

If $H(t)=t^p$, $p\geqq 1$, then $\mathfrak{K}_H^{1/p}$ is the usual metric in $\mathcal{P}$. Copying the proof of Theorem 3, we can obtain the following duality theorem for $\mathfrak{K}_H$.

Corollary 1. *If* $\int_{U_i} H(d(x,a_i))(P_i+Q_i)(dx) < \infty$ *for some* $a_i \in U_i$, $i = 1,\cdots,n$, *then*

$$\begin{aligned}\mathfrak{K}_H(\tilde{P},\tilde{Q}) = \sup\Bigg\{ \sum_{i=1}^n \int_{U_i} f_i\, dp_i + \int_{U_i} g_i\, dP_i \colon f_i, g_i \in \operatorname{Lip}(U_i),\\ i=1,\cdots,n,\ \sum_{i=1}^n f_i(x_i)+g_i(y_i) \leqq \Delta(x,y),\ x,y\in\mathfrak{A}\Bigg\}.\end{aligned} \tag{3.15}$$

If $\tilde{P}$ and $\tilde{Q}$ are sets of tight measures, then the infimum is attained in (3.14).

In the special case $H(t)=t^p$, $\|\mathbf{b}\|_n=|\sum_{i=1}^n b_i|$, we have the following expression for the metric $\mathfrak{K}=\mathfrak{K}_H$.

Corollary 2. *If* $\int_{U_i} d_i(x, a_i)(P_i+Q_i)(dx)<\infty$ *for some* $a_i\in U_i$, $i=1,\cdots,n$, *then*

$$\begin{aligned}\mathfrak{K}(\tilde{P},\tilde{Q})&=\inf\left\{\int_{\mathfrak{A}\times\mathfrak{A}}\sum_{i=1}^n d_i(x_i,y_i)P(d\mathbf{x}\,d\mathbf{y})\colon P\in\mathfrak{P}(\tilde{P},\tilde{Q})\right\}\\ &=\sum_{i=1}^n \sup\left\{\left|\int_{U_i} f\,d(P_i-Q_i)\right|\colon f\in \mathrm{Lip}_{1,1}(U_i)\right\}\\ &=\sum_{i=1}^n \mathscr{A}_{d_i}(P_i,Q_i).\end{aligned}\tag{3.16}$$

The representation (3.16) is a direct consequence of both (3.15) and the one-dimensional Theorem 1.

The distance $\mathfrak{K}_H$ assigns a natural metric structure in $\tilde{\mathscr{P}}$ and an important question is that of finding criteria for the convergence of collections of measures $\tilde{P}\in\tilde{\mathscr{P}}$ in terms of $\mathfrak{K}_H$. Let us state a multi-dimensional analogue of the Kantorovich-Rubinshtein theorem.

Corollary 3. *Let* $\tilde{P}^{(k)}=(P_1^{(k)},\cdots,P_n^{(k)})\in\tilde{\mathscr{P}}$, $k=0,1,\cdots$, *and let*

$$\int_{U_i} H(d(x_i,a_i))P_i^{(k)}(dx_i)<\infty,\qquad i=1,\cdots,n,\ k=0,1,\cdots.$$

Then $\lim_{k\to\infty}\mathfrak{K}_H(\tilde{P}^{(k)},\tilde{P}^{(0)})\to 0$ *if and only if, for all* $i=1,\cdots,n$,

$$P_i^{(m)}\xrightarrow{w}P_i^{(0)}\quad\text{and}\quad\int_{U_i}H(d(x_i,c_i))(P_i^{(m)}-P_i^{(0)})(dx_i)\to 0,\quad m\to\infty,$$

for some (and hence for all) $c_i\in U_i$.

To conclude, we turn our attention to the relationship between Theorem 3 and the multi-dimensional Strassen theorem.

Theorem 4. (See [40].) *Suppose that* (U,d) *is a s.m.s.,*

$$\mathscr{K}\mathscr{F}(\mu)=\inf\{\alpha>0\colon \mu(\mathscr{D}(x)>\alpha)<\alpha\}\tag{3.17}$$

is the Ky Fan functional in $\mathscr{P}_{U^{\times N}}$, *and*

$$\begin{aligned}\Pi(\tilde{P})=\inf\{\alpha>0;\ &P_1(B_1)+\cdots+P_{N-1}(B_{N-1})\\ &\leqq P_N(B_0^\alpha)+\alpha+N-2\ \text{for all}\ B_1,\cdots,B_{N-1}\in\mathfrak{B}(U)\}\end{aligned}\tag{3.18}$$

is the Prokhorov functional in $\mathscr{P}^{\times N}$, *where* $B^\alpha=\{x_N\in U\colon \mathscr{D}(x_1,\cdots,x_N)\leqq\alpha$ *for some* $x_1\in B_1,\cdots,X_{N-1}\in B_{N-1}\}$. *Then*

$$\inf\{\mathscr{K}\mathscr{F}(\mu)\colon \mu\in\mathfrak{P}(\tilde{P})\}=\Pi(\tilde{P}),\tag{3.19}$$

and if $\tilde{P}$ *is a set of tight measures, then the infimum is attained in* (3.17).

This theorem was proved by Schay [130] in the case of a Polish space. From (3.3) and (3.19) follows the next inequality relating the functional $\mathbf{K}(\tilde{P})$ and the Prokhorov functional (3.18) in the multi-dimensional case: for any $H\in\mathscr{H}_1$, $M>0$

and $a \in U$,

$$\Pi(\tilde{P})H(\Pi(\tilde{P})) \leqq \mathbf{K}(\tilde{P}) \leqq H(\Pi(\tilde{P})) + c_1 H(M)\Pi(\tilde{P}) + c_2 \sum_{i=1}^{N} \int_U H(d(x_i, a)) I(d(x_i, a) > N) P_i(dx_i), \tag{3.20}$$

where $c_2 = Q^l$, $l = [\log_2 (A_m N^2)] + 1$, $c_1 = Nc_2$, $[x]$ is the integral part of x and $A_m = \max_{1 \leqq j \leqq m} \{\|(i_1, \cdots, i_m)\|: i_k = 0,\ k \neq j,\ i_j = 1\}$.

We point out that (3.20) easily leads to the assertion in Corollary 3 to Theorem 3.

4. Applications of the Monge–Kantorovich Problem

As was clarified in the preceding sections, the MKP studies natural metrical structure—the Kantorovich functional $\mathcal{A}_c$ and minimal metrics $\mathbf{l}_p$ that have good metrical and topological properties. Thus, for example, the minimal structure of $\mathcal{A}_c$ and $\mathbf{l}_p$ (Theorem 1) is especially useful in problems of stability of stochastic models (see [11]-[20], [141]-[142], [3], [22], [10], [54]). It is natural to use the topological structure of the functionals $\mathcal{A}_c$ and $\mathbf{l}_p$ (Theorem 2) in the limit-type theorems assuring weak convergence plus convergence of moments such as theorems for moments (see [1], [28], [58]) and global limit theorems (see [29]).

In this section, we shall study the Glivenko-Cantelli theorem, a functional limit theorem, and the stability of queueing systems (QS) in terms of the functionals $\mathcal{A}_c$ and $\mathbf{l}_p$. These of course do not exhaust the possible applications of Theorem 1 in the area of stochastics but they do describe a wide spectrum of the possible applications of the theorem.

4.1. Glivenko–Cantelli theorem. Let (U, d) be a s.m.s. and let $\mathcal{P}_U$ be the set of all probability measures on U. Let $X_1, X_2, \cdots$ be a sequence of r.v. with values in U and with respective distributions $P_1, P_2, \cdots \in \mathcal{P}_U$. For any $n \geqq 1$, define the "empirical measure"

$$\mu_n = (\delta_{X_1} + \cdots + \delta_{X_n})/n$$

and "average" measure

$$\bar{P}_n = (P_1 + \cdots + P_n)/n.$$

Let $\mathcal{A}_c$ be the functional (2.1), where $c \in \mathfrak{C}_1$. We now state the well-known theorems of Fortet-Mourier [82], Varadarajan [137], and Wellner [139] in terms of $\mathcal{A}_c$ relying on Theorem 2.

Theorem (Fortet-Mourier). *If $P_1 = P_2 = \cdots = \mu$ and $c_0(x, y) = d(x, y)/(1 + d(x, y))$, then $\mathcal{A}_{c_0}(\mu_n, \mu) \to 0$ almost surely (a.s.) as $n \to \infty$.*

Theorem (Varadarajan). *If $P_1 = P_2 = \cdots = \mu$ and c $(c \in \mathfrak{C}_1)$ is a bounded function, then $\mathcal{A}_c(\mu_n, \mu) \to 0$ a.s. as $n \to \infty$.*

Theorem (Wellner). *If $\bar{P}_1, \bar{P}_2, \cdots$ is a tight sequence, then $\mathcal{A}_{c_0}(\mu_n, \bar{P}_n) \to 0$ a.s. as $n \to \infty$.*

The following theorem extends the results of Fortet-Mourier, Varadarajan and Wellner to the case of an arbitrary functional $\mathcal{A}_c$, $c \in \mathfrak{C}_1$.

Theorem 5. *Suppose that $s_1, s_2 \cdots$ is a sequence of measurable operators in U and*

$$D_i = \sup \{d(s_i x, x) \colon x \in U\},$$

$$L_i = \sup \{d(s_i x, s_i y)/d(x, y) \colon x \neq y,\ x, y \in U\},$$

$$\Theta_i = \min [D_i, (L_i+1)\mathscr{A}_{c_0}(\delta_{X_i}, P_i), 1], \qquad i = 1, 2, \cdots.$$

Let $Y_i = s_i(X_i)$, Q_i be the distribution of Y_i, $\bar{Q}_n = (Q_1 + \cdots + Q_n)/n$ and $\nu_n = (\delta_{Y_1} + \cdots + \delta_{Y_n})/n$. If $\bar{Q}_1, \bar{Q}_2, \cdots$ is a tight sequence,

$$\Theta_n = (\Theta_1 + \cdots + \Theta_n)/n \to 0 \quad a.s., \qquad n \to \infty, \tag{4.1}$$

$c \in \mathfrak{C}_1$ and for some $a \in U$

$$\lim_{M\to\infty} \sup_n \int_U c(x, a) I\{d(x, a) > M\}(\mu_n + \bar{P}_n)(dx) = 0 \ a.s., \tag{4.2}$$

then $\mathscr{A}_c(\mu_n, \bar{P}_n) \to 0$ a.s. as $n \to \infty$.

Proof. From Wellner's theorem (see [139], and also [79], Theorem 8.3) and (2.5), it follows that $\lim_n \mathscr{A}_{c_0}(\nu_n, \bar{Q}_n) = 0$ a.s. We next estimate $\mathscr{A}_{c_0}(\mu_n, P_n)$ obtaining

$$\mathscr{A}_{c_0}(\mu_n, \bar{P}_n) \leqq \mathscr{A}_{c_0}(\nu_n, \bar{Q}_n) + (B_1 + \cdots + B_n)/n, \tag{4.3}$$

where

$$B_i = \sup \left\{ \left| \int_U [f(s_i(x)) - f(x)](\delta_{X_i} - P_i)(dx) \colon f \colon U \to R^1, \right.\right.$$
$$\left. f(x) - |f(y)| \leqq c_0(x, y),\ x, y \in U \right\}.$$

B_i satisfies the estimate $B_i \leqq 2 \min (L_1, 1, 4(L_i+1)\mathscr{A}_{c_0}(\delta_{X_i}, P_i))$. According to (4.1), $\mathscr{A}_{c_0}(\mu_n, \bar{P}_n) \to 0$ a.s. as $n \to \infty$. By Theorem 1 and Strassen's theorem (2.8), it is possible to derive these inequalities: for any positive M,

$$\frac{\pi^2(\mu_n, \bar{P}_n)}{1 + \pi(\mu_n, \bar{P}_n)} \leqq \mathscr{A}_{c_0}(\mu_n, \bar{P}_n) \leqq \pi(\mu_n \bar{P}_n) + \frac{\pi(\mu_n, \bar{P}_n)}{1 + \pi(\mu_n, \bar{P}_n)}, \tag{4.4}$$

and

$$\begin{aligned} \mathscr{A}_c(\mu_n, \bar{P}_n) \leqq\ & H(\pi(\mu_n, \tilde{P}_n)) + 2c_H \pi(\mu_n, \bar{P}_n) H(M) \\ & + c_H \int_U c(x, a) I\{d(x, a) > M\}(\mu_n + \bar{P}_n)(dx), \end{aligned} \tag{4.5}$$

where π is the Lévy-Prokhorov metric. (For a derivation of relations such as (4.4) and (4.5), see [14].) From (4.1), (4.4) and (4.5), it follows that $\mathscr{A}_c(\mu_n, \bar{P}_n) \to 0$ a.s. as $n \to \infty$.

Corollary 1. *If c ($c \in \mathfrak{C}_1$) is a bounded function and $\Theta_n \to 0$ a.s., then $\mathscr{A}_c(\mu_n, \bar{P}_n) \to 0$ a.s. as $n \to \infty$.*

Corollary 1 is a consequence of Wellner's theorem when $s_i(x) = x$, $x \in U$. The following example shows that the conditions imposed in Corollary 1 are actually weaker as compared to the conditions of Wellner's theorem.

EXAMPLE. Let $(U, \|\cdot\|)$ be a separable normed space. Let $x_k \in U$, $\|x_k\| > k^2$, $k = 1, 2, \cdots$, and let $X_k = x_k$ a.s. If we put $s_k(x) = x - x_k$, then $\bar{Q}_n = \delta_0$ and $\Theta_n = 0$ a.s. Thus $\mathscr{A}_c(\mu_n, \bar{P}_n) = 0$ a.s. but $\bar{P}_n$ is not a tight sequence.

In the following we shall assume that $P_1 = P_2 = \cdots = \mu$. In that event, the Glivenko-Cantelli theorem can be stated as follows in terms of $\mathscr{A}_c$ and $\mathbf{l}_p$.

Corollary 2. *Let $c \in \mathfrak{C}_1$ and $\int_U c(x, a)\mu(dx) < \infty$. Then $\mathscr{A}_c(\mu_n, \mu) \to 0$ a.s. as $n \to \infty$. In particular, if*

$$\int_U d^p(x, a)\mu(dx) < \infty, \qquad 0 < p < \infty,$$

then $\mathbf{l}_p(\mu_n, \mu) \to 0$ a.s. as $n \to \infty$.

Corollary 2 was proved for the case $p = 1$ by Fortet and Mourier [82], for $p \geqq 1$ in [116], and for the general case in [117], [43]. Theorem 1 gives an explicit representation of the functionals $\mathscr{A}_c$, $c \in \mathfrak{C}_2$, when $U = R^1$ (see (2.14)). Corollary 2 may be formulated in this case as follows.

Corollary 3. *Let $c \in \mathfrak{C}_2$, $U = R^1$, and $d(x, y) = |x - y|$. Let $F_n(x)$ be the empirical distribution function corresponding to the theoretical distribution function $F(x)$ with $\int c(x, 0)\, dF(x)$ finite. Then*

$$\int_0^1 c(F_n^{-1}(x), F^{-1}(x))\, dx \to 0 \quad \text{a.s.}$$

In particular, if

$$\int |x|^p\, dF(x) < \infty, p \geqq 1,$$

then

$$\int_0^1 |F_n^{-1}(x) - F^{-1}(x)|^p\, dx \to 0 \quad \text{a.s.}$$

Corollary 3 was proved earlier for the cases $p = 1$ in [82]; $p = 2$, $F(x) = x$, $x \in [0, 1]$ in [82]; and $p = 2$, $F(x)$ a continuous strictly increasing function in [127]; for the general case, see [116], [117] and [43].

A consequence of the Fortet-Mourier theorem (see [76], Theorem 7, and [79], Theorem 8.4) is Ranga Rao's result [119]: for any class $\mathscr{M}$ of equicontinuous and uniformly bounded functions,

$$\sup\left\{\left|\int_U f\, d(\mu_n - \mu)\right|; f \in \mathscr{M}\right\} \to 0 \quad \text{a.s.}, \qquad n \to \infty. \tag{4.6}$$

In many problems (see, for example [68] and the references cited there), it is natural to study the closeness of μ_n to μ in terms of the functional

$$\eta_{\mathscr{N}}(\mu_n, \mu) = \sup\left\{\int_U f\, d\mu_n + \int_U g\, d\mu \colon (f, g) \in \mathscr{N}\right\},$$

where $\mathscr{N}$ is some class of pairs of measurable functions on U. Theorem 2 and Corollary 2 to Theorem 5 furnish the convergence of μ_n to μ in terms of $\eta_{\mathscr{N}}$.

Corollary 4. *Let H be a continuous non-negative function such that $\lim_{t\to 0}\sup\{H(s): 0\leqq s\leqq t\}=0$ and there exist a point t_0 and $H_0\in\mathscr{H}_1$ for which $H(t)=H_0(t)$ for $t\geqq t_0$. Let $c=H\circ d$ and $\int_U c(x,a)P(dx)<\infty$. Then $\eta_{\mathcal{N}}(\mu_n,\mu)\to 0$ a.s. as $n\to\infty$, where $\mathcal{N}$ is the class of all pairs of measurable functions (f,g) for which $f(x)+g(y)\leqq c(x,y)$, $x, y\in U$.*

We study next the estimation of the convergence speed in the Glivenko-Cantelli theorem in terms of $\mathscr{A}_c$. Estimates of this sort are useful if one has to estimate not only the speed of convergence of the distribution μ_n to μ in weak metrics but also the speed of convergence of their moments. Thus, for example, if $\mathbf{E}\mathbf{l}_p(\mu_n,\mu)=o(\varphi(n))$, $n\to\infty$, for some $p\in(0,\infty)$, then Theorem 1 and Strassen's theorem (2.8) imply that $(\mathbf{E}(\pi(\mu_n,\mu))^{(p+1)/p'}=o(\varphi(n))$, $n\to\infty$, where $p'=\max(1,p)$ (see [14], [21]) and by Minkowski's inequality it follows that

$$\mathbf{E}\left|\left[\int_U d^p(x,a)\mu_n(dx)\right]^{1/p'}-\left[\int_U d^p(x,a)\mu(dx)\right]^{1/p'}\right|=o(\varphi(n))$$

for any point $a\in U$.

We shall estimate $\mathbf{E}\mathscr{A}_c(\mu_n,\mu)$ in terms of the ε-entropy of the measure μ as was suggested by Dudley in [78] (see also [85]). Let $N(\mu,\varepsilon,\delta)$ be the smallest number of sets of diameter at most 2ε whose union covers U except for a set A_0 with $\mu(A_0)\leqq\delta$. Using Kolmogorov's definition of the ε-entropy of a set U, we call $\log N(\mu,\varepsilon,\varepsilon)$ the *ε-entropy of the measure μ* (see [78]). The next theorem was proved by Dudley [78] for $c=c_0$.

Theorem 6. *Let $c=H\circ d\in\mathfrak{C}_1$ and $H(t)=t^\alpha h(t)$, where $0<\alpha\leqq 1$ and $h(t)$ is a nondecreasing function on $[0,\infty)$. Let $\beta_r=\int_U c^r(x,a)\mu(dx)<\infty$ for some $r>1$ and $a\in U$.*

(a) *If there exist numbers $k\geqq 2$ and $K<\infty$ such that*

$$N(\mu,\varepsilon^{1/\alpha},\varepsilon^{k/(k-2)})\leqq K\varepsilon^{-k}, \tag{4.7}$$

then

$$\mathbf{E}\mathscr{A}_c(\mu_n,\mu)\leqq Cn^{-(1-1/r)/k}, \tag{4.8}$$

where C is a constant depending just on α, k and K.

(b) *If $h(0)>0$ and, for some positive c_1 and δ,*

$$N(\mu,\varepsilon^{1/\alpha},\tfrac{1}{2})\geqq c_1\varepsilon^{-k} \tag{4.9}$$

then there exists a $c_2=c_2(\mu)$ such that

$$\mathbf{E}\mathscr{A}_c(\mu_n,\mu)\geqq c_2 n^{-1/k}. \tag{4.10}$$

The proof of Theorem 6 is based on [78] and the inequality

$$\mathscr{A}_c(\mu,\nu)\leqq 2H(N)\mathscr{A}_{c_\alpha}(\mu,\nu)+2c_H\int c(x,a)I\{d(x,a)>N/2\}(\mu+\nu)(dx), \tag{4.11}$$

where $c_\alpha=d^\alpha/(1+d^\alpha)$, $N>0$ and μ and ν are arbitrary measures on $\mathscr{P}_U$. If $(U,d)=(R^d,\|\cdot\|)$, $m_\gamma=\int\|\mathbf{x}\|^\gamma\mu(dx)<\infty$, where $\gamma=k\alpha d/[(k\alpha-d)(k-2)]$, $k\alpha>d$, $k>2$, then requirement (4.7) is satisfied. If $(U,d)=(R^{k\alpha},\|\cdot\|)$, where $k\alpha$ is an integer and μ is an absolutely continuous distribution, then condition (4.9) is satisfied. The estimate $\mathbf{E}\mathscr{A}_c(\mu_n,\mu)\leqq cn^{-1/k}$ is exact as to order when $h(0)=0$,

$k\alpha$ is an integer, $U = R^{k\alpha}$ and μ is an absolutely continuous distribution having uniformly bounded moments β_r, $r > 1$, and m_γ, $\gamma > 1$.

4.2. Functional central limit theorem. Let $\xi_{n1}, \xi_{n2}, \cdots, \xi_{nk}$, $n = 1, 2, \cdots$, be a double sequence of independent r.v. with d.f. F_{nk}, $k = 1, \cdots, k_n$, obeying the condition of limiting negligibility

$$\lim_{n} \max_{1 \leqq k \leqq k_n} \mathbf{P}(|\xi_{nk}| > \varepsilon) = 0 \tag{4.12}$$

and the conditions

$$\mathbf{E}\xi_{nk} = 0, \qquad \mathbf{E}\xi_{nk}^2 = \sigma_{nk}^2 > 0, \qquad \sum_{k=1}^{k_n} \sigma_{nk}^2 = 0. \tag{4.13}$$

Let $\zeta_{n0} = 0$ and $\xi_{nk} = \xi_{n1} + \cdots + \xi_{nk}$, $1 \leqq k \leqq k_n$, and form a random polygonal line $\zeta_n(t)$ with vertices $(t_{nk}, \mathbf{D}\zeta_{nk})$ (see [36]). Let P_n, from the space of probability measures defined on $(C[0, 1], \|x\| = \sup\{|x(t)|: t \in [0, 1]\})$ be the distribution of $\zeta_n(t)$ and let W be a Wiener measure in $C[0, 1]$. On the basis of Theorem 2, we can state the following version of the Donsker–Prokhorov theorem for the functional $\mathscr{A}_c$.

Theorem 7. *Suppose that conditions* (4.12) *and* (4.13) *hold and that* $\mathbf{E}H(|\xi_{nk}|) < \infty$, $k = 1, 2, \cdots, k_n$, $n = 1, 2, \cdots, H \in \mathscr{H}_1$. *Then the convergence* $\mathscr{A}_c(P_n, W) \to 0$, $n \to \infty$, *is equivalent to the fulfillment of the Lindeberg condition*

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} \int_{|x|>\varepsilon} x^2 \, dF_{nk}(x) = 0, \qquad \varepsilon > 0, \tag{4.14}$$

and the Bernshtein condition

$$\lim_{N\to\infty} \limsup_{n\to\infty} \sum_{k=1}^{k_n} \int_{|x|>N} H(|x|) \, dF_{nk}(x) = 0. \tag{4.15}$$

Proof. By Theorem 3.1 of [36], the necessity of (4.14) is a straightforward consequence of Theorem 2. Let us prove the necessity of (4.15). Define the functional $b: C[0, 1] \to R^1$ by the relation $b(x) = x(1)$. Since, for any $N > 2\sqrt{2}$,

$$\int_N^\infty \mathbf{P}(\|\zeta_n\| > t) \, dH(t) \leqq 4c_H \int_{N/2}^\infty \mathbf{P}(|\zeta_{n,k_n}| \geqq t) \, dH(t),$$

it follows that $\mathbf{E}H(\|\zeta_n\|) < \infty$ for all $n = 1, 2, \cdots$. By Theorem 2, the relations $P_n \xrightarrow{w} W$ and $\int H(\|x\|)(P_n - W)(dx) \to 0$ hold as $n \to \infty$ and since for any N

$$\mathbf{E}H(|b(\zeta_n)|)I\{|b(\zeta_n)| > N\} \leqq 2 \int_{N_1}^\infty \mathbf{P}(\|\zeta_n\| > t) \, dH(t), \qquad N_1 = H^{-1}(H(N)/2),$$

we have $P_n \circ b^{-1} \xrightarrow{w} W \circ b^{-1}$ and $\int H(\|x\|)(P_n \circ b^{-1} - W \circ b^{-1})(dx) \to 0$ as $n \to \infty$. By virtue of Theorem 1 of [28], the necessity of condition (4.15) has been proved. The sufficiency of (4.14) and (4.15) is proved in a similar way.

Corollary 1 (Bernshtein). *Let* $\xi_1, \xi_2, \cdots$ *be a sequence of independent r.v. such that* $\mathbf{E}\xi_i^2 = b_i$ *and* $\mathbf{E}|\xi_i|^p < \infty$, $i = 1, 2, \cdots, p > 2$. *Let* $B_n = b_1 + \cdots + b_n$, $\zeta_n = \xi_1 + \cdots + \xi_n$ *and let the sequence* $B_n^{-1/2}\xi_j$, $j = 1, 2, \cdots$, *satisfy the limiting negligibil-*

ity condition. Let $X_n(t)$ be the random polygonal line with vertices $(B_k/B_n, B_n^{-1/2}\zeta_k)$ and let P_n be its distribution. Then the convergence

$$(4.16)\qquad \mathbf{1}_p(P_n, W)\to 0, \qquad n\to\infty,$$

is equivalent to the fulfillment of the condition

$$(4.17)\qquad \lim_{n\to\infty} B_n^{-p/2}\sum_{i=1}^{n}\mathbf{E}|\xi_j|^p=0.$$

The following statement may be called the Bernshtein-Kantorovich invariance principle.

Corollary 2. *Suppose that $c, c'\in\mathfrak{C}_1$, the series scheme $\{\xi_{nk}\}$ satisfies the conditions of Theorem* 7, *and conditions* (4.14) *and* (4.15) *hold. Then $\mathscr{A}_{c'}(P_n\circ b^{-1}, W\circ b^{-1})\to 0$ as $n\to\infty$ for any functional on $C[0,1]$ for which $N(b; c, c')=$* $\sup\{c'(b(x), b(y)/c(x,y)\colon x\neq y,\ x, y\in C[0,1]\}<\infty$.

Let $c'(t,s)=H'(|t-s|)$ and $t, s\in R^1$. Consider the following examples of functionals b with finite $N(b; c, c')$.

(a) *If $H=H'$ and b has a finite Lipschitz form*

$$\|b\|_L=\sup\{|b(x)-b(y)|/\|x-y\|\colon x\neq y,\ x, y\in C[0,1]\}<\infty,$$

then $N(b; c, c')<\infty$. Functionals such as these are $b_1(x)=x(a)$, $a\in[0,1]$; $b_2(x)=$ $\max\{x(t)\colon t\in[0,1]\}$; $b_3(x)=\|x\|$ *and* $b_4(x)=\int_0^1\varphi(x(t))\,dt$, *where* $\|\varphi\|_L\leqq 1$.

(b) *Let $H(f)=t^p$ and $H'(t)=t^{p'}$, $0<p<p'$. Then $N(b_3^{p/p'}; c, c')<\infty$ and $N(b_4; c, c')<\infty$ if*

$$|\varphi(x)-\varphi(y)|\leqq\|x-y\|^{p/p'}, \qquad x, y\in C[0,1].$$

We state one further consequence of Theorem 7.

Let the series scheme $\{\xi_{nk}\}$ satisfy the conditions of Theorem 7 and let $\eta_n(t)=\zeta_{nk}$ for $t\in(t_{n(k-1)}, t_{nk})$, $k=1,\cdots,k_n$, $\eta_n(0)=0$. Let $\hat{P}_n$ be the distribution of η_n. The distribution $\hat{P}_n$ belongs to the space of probability measures defined on the Skorokhod space $D[0,1]$ (see [2]).

Corollary 3. *The convergence $\mathscr{A}_c(\hat{P}_n, W)\to 0$ as $n\to\infty$ is equivalent to the fulfillment of* (4.14) *and* (4.15).

4.3. Stability of queueing systems. As a model example of the applicability of Kantorovich's theorem in the stability problem for queueing systems, we consider the stability of the system $G|G|1|\infty$. (A detailed discussion of this problem and its generalizations is presented in Kalashnikov and Rachev's paper: "A characterization of queueing system models and its stability," which will be published in the collection: Stability of Stochastic Models, VNIISI, Moscow, 1984.) Sequences of non-negative r.v. $\{e_n\}_{n=0}^{\infty}$ and $\{s_n\}_{n=0}^{\infty}$ are treated as sequences of the lengths of the time intervals between the n-th and $(n+1)$-st arrivals and the service times of the n-th arrival, respectively. Define the recursive sequence

$$(4.18)\qquad w_0=0, \qquad w_{n+1}=\max(w_n+s_n-e_n, 0), \qquad n=1,2,\cdots.$$

The quantity w_n may be viewed as the waiting time for the n-th arrival to begin to be serviced. Introduce the notation: $e_{j,k}=(e_j,\cdots,e_k)$, $s_{j,k}=(s_j,\cdots,s_k)$, $k>j$,

$\mathbf{e}=(e_0, e_1, \cdots)$, and $\mathbf{s}=(s_0, s_1, \cdots)$. Along with the model defined by relations (4.18), we consider a sequence of analogous models by indexing it with the letter r ($r \geqq 1$). Namely, all quantities pertaining to the r-th model will be designated in the same way as the model (4.18) but will have a superscript r: $e_n^{(r)}$, $s_n^{(r)}$, $w_n^{(r)}$, and so on. It is convenient to regard the value $r=\infty$ (which can be omitted) as corresponding to the original model. All of the random variables are assumed to be defined on the same probability space. For brevity, functionals Φ depending just on the distributions of the r.v. X and Y will be denoted by $\Phi(X, Y)$.

For the system $G|G|1|\infty$ in question, define for $k \geqq 1$ non-negative functions φ_k on $(R^k, \|\mathbf{x}\|)$, $\|(x_1, \cdots, x_k)\| = |x_1| + \cdots + |x_k|$, as follows:

$$\varphi_k(\xi_1, \cdots, \xi_k, \eta_1, \cdots, \eta_k) = \max[0, \eta_k - \xi_k, (\eta_k - \xi_k) + (\eta_{k-1} - \xi_{k-1}), \cdots, (\eta_k - \xi_k) + \cdots + (\eta_1 - \xi_1)].$$

It is not hard to see that $\varphi_k(e_{n-k,n-1}, s_{n-k,n-1})$ is the waiting time for th n-th arrival under the condition that $w_{n-k}=0$.

Let $c \in \mathfrak{C}_1$. The system $G|G|1|\infty$ is uniformly stable with respect to the functional $\mathscr{A}_c$ on finite time intervals if, for every positive T, the following limit relation holds:

$$\delta^{(r)}(T; \mathscr{A}_c) \equiv \sup_{n \geqq 0} \max_{1 \leqq k \leqq T} \mathscr{A}_c[\varphi_k(e_{n,n+k-1}, s_{n,n+k-1}), \varphi_k(e^{(e)}_{n,n+k-1}, s^{(r)}_{n,n+k-1})] \to 0 \quad \text{as } r \to \infty. \tag{4.19}$$

The relation (4.19) means that the largest deviation between the variables w_{n+k} and $w^{(r)}_{n+k}$, $k=1, \cdots, T$, converges to zero as $r \to \infty$ if at time n both compared systems are free, and for any positive T this convergence is uniform in n.

Theorem 8. *If for each $r=1, 2, \cdots, \infty$ the sequences $\mathbf{e}^{(r)}$ and $\mathbf{s}^{(r)}$ are independent, then*

$$\delta_c^{(r)}(T; \mathscr{A}_c) \leqq c_H \sup_{n \geqq 0} \mathscr{A}_c(e_{n,n+T-1}, e^{(r)}_{n,n+T-1}) + c_H \sup_{n \geqq 0} \mathscr{A}_c(s_{n,n+T-1}, s^{(r)}_{n,n+T-1}). \tag{4.20}$$

Corollary 1. *Under the assumptions of Theorem 8, for any $p \in [0, \infty]$,*

$$\delta_c^{(r)}(T; \mathbf{1}_p) \leqq \sup_{n \geqq 0} \mathbf{1}_p(e_{n,n+T-1}, e^{(r)}_{n,n+T-1}) + \sup_{n \geqq 0} \mathbf{1}_p(s_{n,n+T-1}, s^{(r)}_{n,n+T-1}). \tag{4.21}$$

From (4.20) and (4.21), it is possible to derive an estimate of the stability of the system $G|G|1|\infty$ in the sense of (4.19). It can be expressed in terms of the deviations of the vectors $e^{(r)}_{n,n+T-1}$ and $s^{(r)}_{n,n+T-1}$ from $e_{n,n+T-1}$ and $s_{n,n+T-1}$, respectively. Such deviations are easy to estimate if we impose additional restrictions on $\mathbf{e}^{(r)}$ and $\mathbf{s}^{(r)}$, $r=1, 2, \cdots$. For example, when the terms of the sequences are independent, the following estimates hold:

$$\mathscr{A}_c(e_{n,n+T-1}, e^{(r)}_{n,n+T-1}) \leqq c_H^q \sum_{j=n}^{n+T-1} \mathscr{A}_c(e_j, e_j^{(r)}), \qquad q = [\log_2 T] + 1,$$

$$\mathbf{1}_p(e_{n,n+T-1}, e^{(r)}_{n,n+T-1}) \leqq \sum_{j=n}^{n+T-1} \mathbf{1}_p(e_j, e_j^{(r)}), \qquad 0 \leqq p \leqq \infty,$$

which can be even further simplified when the terms of these sequences are identically distributed. On the basis of (4.20) and (4.21), it is possible to construct stability estimates for the system that are uniform over the entire time axis (see [11], [12], [15], [22], [99], [3]).

To conclude, the author considers it his accepted duty to thank Academician Yu. V. Prokhorov for suggesting that this review paper be written. The author offers his thanks to Academician L. V. Kantorovich and to R. L. Dobrushin, V. M. Zolotarev and V. N. Sudakov for stimulating discussions relating to the content of the article. The author expresses his thanks to the governing body of the Mathematics Institute of the USSR Academy of Sciences and the Probability-Theoretic Section for the beautiful working conditions allotted to him.

Received by the editors
May 18, 1984

REFERENCES[1]

[1] S. N. BERNSHTEIN, *Collected Works*, Vol. 4, Nauka, Moscow, 1969. (In Russian.)
[2] P. BILLINGSLEY, *Convergence of Probability Measures*, John Wiley, New York, 1968.
[3] A. A. BOROVKOV, *Asymptotic Methods in Queueing Theory*, John Wiley, New York, 1984.
[4] S. S. VALLENDER, *Calculation of the Wasserstein distance between probability distributions on the line.* Theory Prob. Appl., 18 (1973), pp. 784-786.
[5] L. N. WASSERSTEIN, *Markov processes over denumerable products of spaces describing large systems of automata*, Problems of Information Transmission, 5, 3 (1969), pp. 47-52.
[6] A. M. VERSHIK, *Some remarks on infinite-dimensional linear programming problems*, Uspekhi Mat. Nauk, 25, 5 (1970), pp. 117-124. (In Russian.)
[7] M. YA. VYGODSKII, *The origins of differential geometry.* In: G. Monge, *Applications of Analysis to Geometry*, Main Editorial Staff of General Technical Disciplines, Moscow-Lengingrad, 1936. (In Russian.)
[8] R. L. DOBRUSHIN, *Gibbsian random fields for lattice systems with pairwise interaction*, Funct. Anal. Appl., 2, 4 (1968), pp. 292-301.
[9] R. L. DOBRUSHIN, *Uniqueness problem for a Gibbsian random field and the problem of phase transitions*, Funct. Anal. Appl., 2, 4 (1968), pp. 302-312.
[10] R. L. DOBRUSHIN, *Prescribing a system of random variables by conditional distributions*, Theory Prob. Appl., 15 (1970), pp. 458-486.
[11] V. M. ZOLOTAREV, *On the continuity of stochastic sequences generated by recurrence procedures*, Theory Prob. Appl., 20 (1975), pp. 819-832.
[12] V. M. ZOLOTAREV, *Qualitative estimates in problems of continuity of queuing systems*, Theory Prob. Appl., 20 (1975), pp. 211-213.
[13] V. M. ZOLOTAREV, *The effect of stability for a characterization of distributions*, Zap. Nauchn. Sem. LOMI, 61 (1976), pp. 38-55. (In Russian.)
[14] V. M. ZOLOTAREV, *Metric distances in spaces of random variables and their distributions*, Math. USSR Sb., 30, 3 (1976), pp. 373-401.
[15] V. M. ZOLOTAREV, *On stochastic continuity of queueing systems of type* $G|G|1$, Theory Prob. Appl., 21 (1976), pp. 250-269.
[16] V. M. ZOLOTAREV, *Approximation of distributions of sums of independent random variables with values in infinite-dimensional spaces*, Theory Prob. Appl., 21 (1976), pp. 721-737.
[17] V. M. ZOLOTAREV, *Ideal metrics in the problem of approximating distributions of sums of independent random variables*, Theory Prob. Appl., 22 (1977), pp. 433-449.
[18] V. M. ZOLOTAREV, *On pseudomoments*, Theory Prob. Appl., 23 (1978), pp. 269-278.

[1] The author would like to thank R. L. Dobrushin, V. L. Levin and R. M. Shortt for supplementing the original list of references with the titles [143]-[155].

[19] V. M. ZOLOTAREV, *On properties and relationships between certain types of metrics*, Zap. Nauchn. Sem. LOMI, 87 (1979), pp. 18-35. (In Russian.)

[20] V. M. ZOLOTAREV, *Probability metrics*, Theory Prob. Appl., 28 (1983), pp. 278-302.

[21] TS. G. IGNATOV AND S. T. RACHEV, *Minimal ideal probability metrics*. In: Stability of Stochastic Models, VNIISI, Moscow, 1983, pp. 36-48. (In Russian.)

[22] V. V. KALASHNIKOV, *Quantitative Analysis of the Behavior of Complex Systems by the Method of Test Functions*, Nauka, Moscow, 1978. (In Russian.)

[23] L. V. KANTOROVICH, *On the transfer of masses*, Dokl. Akad. Nauk. SSSR, 37, 7-8 (1942), pp. 227-229. (In Russian.)

[24] L. V. KANTOROVICH, *On a problem of Monge*, Uspekhi Mat. Nauk, 3, 2 (1948), pp. 225-226. (In Russian.)

[25] L. V. KANTOROVICH AND G. P. AKILOV, *Functional Analysis*, Nauka, Moscow, 1977. (In Russian.)

[26] L. V. KANTOROVICH AND G. SH. RUBINSHTEIN, *On a function space and some extremum problems*, Dokl. Akad. Nauk. SSSR, 115, 6 (1957), pp. 1058-1061. (In Russian.)

[27] L. V. KANTOROVICH AND G. SH. RUBINSHTEIN, *On the space of completely additive functions*, Vestnik LGU, Ser. Mat., Mekh. i Astron., 7, 2 (1958), pp. 52-59. (In Russian.)

[28] V. M. KRUGLOV, *Convergence of numerical characteristics of independent random variables with values in a Hilbert space*, Theory Prob. Appl., 18 (1973), pp. 694-712.

[29] V. M. KRUGLOV, *Global limit theorems*, Zap. Nauchn. Sem. LOMI, 61 (1976), pp. 84-101. (In Russian.)

[30] V. L. LEVIN, *Duality and approximation in the problem of mass transfer*. In: Mathematical Economics and Functional Analysis, Nauka, Moscow, 1974, pp. 94-108. (In Russian.)

[31] V. L. LEVIN, *On the problem of mass transfer*, Soviet Math. Dokl., 16, 5 (1975), pp. 1349-1353.

[32] V. L. LEVIN, *The Monge-Kantorovich problem on mass transfer*. In: Methods of Functional Analysis in Mathematical Economics, Nauka, Moscow, 1978, pp. 23-55. (In Russian.)

[33] V. L. LEVIN, *On duality theorems in the Monge-Kantorovich problem*, Uspekhi Mat. Nauk, 32, 3 (1977), pp. 171-172. (In Russian.)

[34] V. L. LEVIN AND A. A. MALYUTIN, *The mass transfer problem with discontinuous cost function and a mass setting for the problem of duality of convex extremum problems*, Uspekhi Mat. Nauk, 34, 3 (1979), pp. 3-68. (In Russian.)

[35] G. D. MAKAROV, *Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed*, Theory Prob. Appl., 26 (1981), pp. 803-806.

[36] YU. V. PROKHOROV, *Convergence of random processes and limit theorems in probability theory*, Theory Prob. Appl., 1 (1956), 157-214.

[37] S. T. RACHEV, *On a Hausdorff metric construction in the space of probability measures*, Zap. Nauchn. Sem. LOMI, 87 (1979), pp. 87-103. (In Russian.)

[38] S. T. RACHEV, *On minimal metrics in the space of real-valued random variables*, Soviet Dokl. Math., 23, 2 (1981), pp. 425-432.

[39] S. T. RACHEV, *Hausdorff metric construction in the probability measure space*, Pliska Studia Mathem. Bulgarica, 7 (1984), pp. 152-162.

[40] S. T. RACHEV, *Minimal metrics in the space of real random variables with fixed one-dimensional marginal distributions*, Tr. VNIISI, Moscow, 1981, pp. 112-128. (In Russian.)

[41] S. T. RACHEV AND G. S. CHOBANOV, *Existence and uniqueness on a boundary Gibbsian distribution. Dobrushin-Sinai approach*. In: Lectures on Stochastic Problems in Contemporary Physics, Izd-vo Sof. Univ., Sofia, 1983, pp. 42-60. (In Bulgarian.)

[42] S. T. RACHEV, *On the ζ-structure of mean and uniform distances*, Soviet Math. Dokl., 30, 2 (1984), pp. 369-372.

[43] S. T. RACHEV, *On a class of minimal functionals on a space of probability measures*, Theory Prob. Appl., 29 (1984), pp. 41-49.

[44] S. T. RACHEV, *On a problem of Dudley*, Soviet Math. Dokl., 29, 2 (1984), pp. 162-164.

[45] S. T. RACHEV, *Compactness in probability measure space*. In: Proceedings of the Third European Young Statisticians Meeting. Ed. by H. Caeyers et al., Katholieke Universiteit, Leuven, 1983, pp. 148-152.

[46] G. SH. RUBINSHTEIN, *Duality in mathematical programming and some questions in convex analysis*, Uspekhi Mat. Nauk, 25, 5 (1970), pp. 171-201. (In Russian.)

[47] A. V. SKOROKHOD, *Limit theorems for stochastic processes*, Theory Prob. Appl., 1 (1956), pp. 261-290.

[48] V. N. SUDAKOV, *Measures on subsets of direct products*, Zap. Nauchn. Sem. LOMI, 29 (1972), pp. 113-128. (In Russian.)

[49] V. N. SUDAKOV, *On the existence of bistochastic densities*, Zap. Nauchn. Sem. LOMI, 29 (1972), pp. 129-146. (In Russian.)

[50] V. N. SUDAKOV, *Geometric problems in the theory of infinite-dimensional probability distributions*, Tr. Mat. Inst. V. A. Steklov, Akad. Nauk. SSSR, 141 (1976). (In Russian.)

[51] A. L. TOOM, *A family of networks of formal neutrons*, Soviet Math. Dokl., 9, 6 (1968), pp. 1338-1344.

[52] G. H. HARDY, J. E. LITTLEWOOD AND G. POLYA, *Inequalities*, Cambridge Univ. Press, London, 1951.

[53] P. L. HENNEQUIN AND A. TORTRAT, *Théorie des probabilités et quelques applications*, Masson, Paris, 1965.

[54] P. J. HUBER, *Robust Statistical Procedures*, CBMS Regional Conference Series in Applied Mathematics, 27, Society for Industrial and Applied Mathematics, Philadelphia, 1977.

[55] M. G. SHNIRMAN, *On the question of ergodicity of a Markov chain*, Problemy Kib., 20 (1980), pp. 125-222. (In Russian.)

[56] I. A. SHOLPO, *ε-minimal metrics*, Theory Prob. Appl., 28 (1983), pp. 854-855.

[57] A. DE ACOSTA, *Invariance principles in probability for triangle arrays of B-valued random vectors and some applications*, Ann. Probab., 10 (1982), pp. 346-373.

[58] A. ACOSTA AND G. GINE, *Convergence of moments and related functionals in the general central limit theorem in Banach spaces*, Z. Wahrsch. Verw. Geb., 48, 2 (1979), pp. 213-231.

[59] P. APPELL, *Mémoires sur les dèblais et remblais des systèmes continues et discontinues*, Memoires des sav. étr., 2 sér., 1884, 29, 3 mémoire.

[60] P. APPELL, *Le problème géométrique des déblais et remblais.* In: memorial des Sci. Math., Paris.: 1928, Fasc. 27, pp. 1-34.

[61] B. BALDESSARI, *Metodi di quantificazione basati sigli indici di dissomiglianza e di inequalianza*, Metron, XXXII (1974), pp. 197-219.

[62] P. BARTFAI, *Über die Entfernung der Irrahrtswegge*, Studia Sci. Math. Hungaria, V (1970), pp. 41-49.

[63] J. BASS, *Sur la compatibilité des functions de répartition*, C.R. Acad. Sci. Paris, ser. A, 240 (1955), pp. 839-841.

[64] S. BERTINO, *Una generalizzazione della dissomiglianza di C. Gini tra variabili casuali semplici*, Giorn. Ist. Italiano Attuari, 29 (1966), pp. 155-178.

[65] S. BERTINO, *Gli indici di dissomiglianza e la stima dei parametri.* In. Studi di probabilità, statistica e ricerca operativa in onore di Gieseppe Pomlily, Oderisi, Gubbio, 1971.

[66] S. BERTINO, *Sulla media e la varianza dell'indice semplice di dissomiglianza nel caso di campioni provenienti da una stessa variabile casuale assolutemante continue*, Metron, XXX (1972), pp. 256-281.

[67] S. CAMBANIS AND G. SIMONS, *Probability and expectation inequalities*, Z. Wahrsch. Verw. Geb., 59, 1 (1982), pp. 1-25.

[68] S. CAMBANIS, G. SIMONS AND W. STOUT, *Inequalities for* $\mathrm{E}k(X, Y)$ *when the marginals are fixed*, Z. Wahrsch. Verw. Geb., 36, 4 (1976), pp. 285-294.

[69] M. CASSANDO, E. OLIVIERI, A. PELLEGRINOTTI AND E. PRESUTT, *Existence and uniqueness of DLR measures for unbounded spin systems*, Z. Wahrsch. Verw. Geb., 41, 3 (1978), pp. 313-334.

[70] G. DALL'AGLIO, *Sugli estremi dei momenti delle funzioni di ripartizione dopia*, Ann. Scuola Normale Superiore Di Pisa, ser. Cl. Sci., 3, 1 (1956), pp. 33-74.

[71] G. DALL'AGLIO, *Les fonctions extrêmes de la classe de Frêchet à* 3 *dimensions*, Publ. Inst. Stat. Univ. Paris, IX (1960), pp. 175-188.

[72] G. DALL'AGLIO, *Fréchet classes and compatibility of distribution functions*, Symp. Math., 9 (1972), pp. 131-150.

[73] C. D. DOWSON AND B. U. LANDAU, *The Fréchet distance between multivariate normal distributions*, J. Multivar. Anal., 12 (1982), pp. 450-455.

[74] R. G. DOUGLAS, *On extremal measures and subspace density*, Michigan Math. J., 11 (1964), pp. 243-246.

[75] R. G. DOUGLAS, *On extremal measures and subspace density*, II, Proc. Amer. Math. Soc., 17 (1976), pp. 1363–1365.

[76] R. M. DUDLEY, *Convergence of Baire measures*, Studia Math., XXVII, 3 (1966), pp. 251–268.

[77] R. M. DUDLEY, *Distances of probability measures and random variables*, Ann. Math. Statist., 39 (1968), pp. 1563–1572.

[78] R. M. DUDLEY, *The speed of mean Glivenko-Cantelli convergence*, Ann. Math. Statist., 40, (1969), pp. 40–50.

[79] R. M. DUDLEY, *Probability and metrics*, Aarhus Univ: Aarhus, 1976.

[80] X. FERNIQUE, *Sur le théorème de Kantorovich-Rubinstein dans les espaces polonais*, Lecture Notes Mathematics, 850, Springer, Berlin, 1981, pp. 6–10.

[81] A. FORCINA AND G. GALMACEI, *Sulla distribuzione dell'indice semplice di dissomiglianza*, Metron, XXII (1974), pp. 361–378.

[82] R. FORTET AND B. MOURIER, *Convergence de la repartition empirique vers la repartition théoretique*, Ann. Sci. Ecole Norm. Sup., 70, 3 (1953), pp. 267–285.

[83] M. FRÈCHET, *Sur les tableaux de corrélation dont les marges sont données*, Ann. Univ. de Lyon, 14 (1951), pp. 53–77.

[84] M. FRÉCHET, *Sur la distance de deux lois de probabilité*, Publ. Inst. Statist. Univ. Paris, 6, 2 (1957), pp. 185–198.

[85] P. GÄNSSLER, *A note on a result of Dudley on the speed of Glivenko-Cantelli convergence*, Ann. Math. Statist., 41 (1970), pp. 1339–1343.

[86] U. GARSIA-PALOMARES AND E. M. GINE, *On the linear programming approach to the optimality property of Prokhorov's distance*, J. Math. Anal. Appl., 60 (1977), pp. 596–600.

[87] A. DE S. GERMAINE, *Etudes sur le problème des déblais et remblais*, Extrait des Mémoires de l'Academie Nationale de Caen, 1886.

[88] C. GINI, *Di una misura delle relazioni tra le graduatorie di due caratteri*, Appendix to: A. Hancini, *Le Elezioni Generali Politiche de* 1913 *nel comune di Roma*, Ludovico Cecehini, Rome, 1914.

[89] C. GINI, *Di una misura della dissomiglianza tra due gruppi di quatita e della sue applicazioni allo studio delle relazioni statistiche*, Atti de Reale Istituto Veneto di Sci., Lettera ed Arti, LXXIV (1914–1915), pp. 185–213.

[90] C. GINI, *Sul criterio di concordanza tra due caratteri*, Atti del Reale Instituto Veneto di Sci., Lettera ed Arti, LXXV (1915–1916), pp. 309–331.

[91] C. GINI, *Corso di statistica*, Veschi, Roma, 1951.

[92] C. GINI, *La dissomiglianza*, Metron, XXIV (1965), pp. 85–215.

[93] C. R. GIVENS AND R. M. SHORTT, *A class of Wasserstein metrics for probability distributions*, Manuscript, Dept. Math. Comp. Sci. Michigan Tech. Univ., Houghton, MI, 1984.

[94] R. M. GRAY, D. L. NEUHOFF AND P. C. SHIELDS, *A generalization to Ornstein's $\bar{d}$-distance with applications to information theory*, Ann. Probab., 3 (1975), pp. 315–328.

[95] R. M. GRAY AND D. S. ORNSTEIN, *Block coding for discrete stationary $\bar{d}$-continuous channels*, IEEE Trans. Inform. Theory, IT, 25 (1979), pp. 292–306.

[96] R. M. GRAY, D. S. ORNSTEIN AND R. L. DOBRUSHIN, *Block synchronization, sliding-block coding, invulnerable sources and zero error codes for discrete noisy channels*, Ann. Probab., 8 (1980), pp. 639–674.

[97] A. HERZEL, *Il valore medio e la varianza dell'indice semplice di dissomiglianza negli universi dei campioni bernouliano ed esaustivo.* In: Risultati preliminari della teoria dei picooli campioni (Biblioteka del "Metron," serie C: Note e Commenti vol.), Roma, 1963.

[98] W. HOEFFDING, *Masstabvariate korrelationstheorie*, Sehr. Math. Inst. Univ. Berlin, 5 (1950), pp. 181–233.

[99] V. V. KALASHNIKOV AND S. A. ANICHKIN, *Continuity of random sequences and approximation of Markov chains*, Adv. Appl. Probab., 13 (1981), pp. 402–414.

[100] J. M. KEMPERMAN, *On the role of duality in the theory of moments. Semi-infinite programming and applications*, Lecture Notes in Economic and Mathematical Systems, 215, Springer, Berlin, 1983, pp. 63–92.

[101] W. M. KRUSKAL, *Ordinal measures of association*, J. Amer. Statist. Assoc., 53 (1958), pp. 814–861.

[102] T. L. LAI AND M. ROBBINS, *Maximally dependent random variables*, Proc. Nat. Acad. Sci. USA, 73 (1976), pp. 286–288.

[103] G. Landenna, *La dissomiglianza*, Statistica, 1956, anno XVI, pp. 21-57.
[104] G. Landenna, *L'indice semplice di dissomiglianza tra campioni ottenuti con estrazione in blocco*, Statistica, 1961, anno XXI H., Luglio settembre.
[105] G. Leti, *Il termine generico delle tabelle di congraduazione e di contrograduazione*, Biblioteka del Metron, 1962, pp. 1-27.
[106] G. Leti, *Una misura della mutabilità dell'universo dei campioni*, Atti della XXII Riunione Scientifica della Società Italiana di Statistica, Roma 27-28. Ottobre 1962, Roma, 1963.
[107] E. L. Lehman, *Some concepts of dependence*, Ann. Math. Statist., 37 (1966), pp. 1137-1153.
[108] P. Lévy, *Distance de deux variables aléatoires et distance de deux de probabilité.* In: Generalités sur les probabilités, Fréchet, ed., Gauthiers-Villars, Paris, 1950.
[109] G. G. Lorentz, *An inequality for rearrangements*, Amer. Math. Monthly, 60 (1953), pp. 176-179.
[110] L. Lucia, *Variabilita superficiale e dissomiglianza tra distribuzioni semplici*, Metron, XXIV (1965), pp. 225-331.
[111] C. L. Mallows, *A note on asymptotic joint normality*, Ann. Math. Statist., 43 (1972), pp. 508-515.
[112] G. Monge, *Mémoire sur la théorie des déblais et des remblais*, Histoire de l'Académie des sciences de Paris, avec les Mémoires de mathématique et de physique pour la même année, 1781, pp. 666-704, avec 2 pl.
[113] I. Olkin and F. Pukelheim, *The distance between two random vectors with given dispersion matrices*, Linear Algebra and Appl., 48 (1982), pp. 257-263.
[114] D. S. Ornstein, *Entropy is enough to classify Bernoulli shifts but not K-automorphisms*, In the collection Congress International Mathématiciens, 1982, vol. 2, Nauka, Moscow, 1971, pp. 571-575.
[115] G. Pompily, *La mutabilità dell'universo dei campioni, nel volume*, Scritti matematici in onore di Filippo Sibirani, Bologna, 1956.
[116] S. T. Rachev, *Minimal metrics in the random variables space.* In: Probability and Statistical Inference Proceedings of the 2nd Pannonian Symp., W. Grossmann et al., eds., D. Reidel Company, Dordrecht, 1982, pp. 319-327.
[117] S. T. Rachev, *Minimal metrics in the random variable space*, Pub. Inst. Statist. Univ. Paris, XXVII (1982), pp. 27-47.
[118] S. T. Rachev, *Minimal metrics in the real valued random variables space*, Lecture Notes in Mathematics, 982, Springer, Berlin, 1983, pp. 172-190.
[119] R. R. Rao, *Relations between weak and uniform convergence of measures with applications*, Ann. Math. Statist., 33 (1962), pp. 659-680.
[120] Y. Rinott, *Multivariate majorization and rearrangement inequalities with some applications to probability and statistics*, Israel J. Math., 45 (1973), pp. 60-77.
[121] L. Rushendorff, *Inequalities for expectation of Δ-monotone functions*, Z. Wahroch., Verw. Geb., 54 (1980), pp. 341-349.
[122] T. Salvemini, *Sigli indici di omofilia*, Atti 1ª Riunione della Soc. Ital. di Statistica, Pisa, 1939.
[123] T. Salvemini, *Sul calcolo degli indici di concordanza tra due caratteri quantitativi*, Atti della VI Riunione della Soc. Ital. di Statistica, Roma, 1943.
[124] T. Salvemini, *Nuovi procedimenti per il calcolo degli indici di dissomiglianza e di connesione*, Statistica, 1949, anno IX, pp. 3-36.
[125] T. Salvemini, *L'indice di dissomiglianza fra distributioni continua*, Metron, 16 (1957), pp. 75-100.
[126] T. Salvemini, *L'indice quadratico di dissomiglianza fra distribuzioni gaussiane*, Atti della XIII e XIV Riunione Scientifica, Societa Italiana di Statistica, 1953-1954, pp. 297-301.
[127] E. Samuel and R. Bachi, *Measures of the distance of distribution functions and some applications*, Metron, XXIII (1964), pp. 83-122.
[128] M. Schaeffer, *Note on the k-dimensional Jensen inequality*, Ann. Probab., 4 (1976), pp. 502-504.
[129] G. Schay, *Nearest random variables with given distributions*, Ann. Probab., 2 (1974), pp. 163-166.
[130] G. Schay, *Optimal joint distributions of several random variables with given marginals*, Stud. Appl. Math., LXI (1979), pp. 179-183.
[131] S. Sherman, *On a conjecture concerning doubly stochastic matrices*, Proc. Amer. Math. Soc., 3 (1952), pp. 511-513.

[132] R. M. SHORTT, *Strassen's marginal problems in two or more dimensions*, Z. Wahrsch. Verw. Geb., 64 (1983), pp. 313-325.

[133] V. STRASSEN, *The existence of probability measures with given marginals*, Ann. Math. Statist., 36 (1965), pp. 423-439.

[134] A. SZULGA, *On the Wasserstein metric.* In: Transactions of the 8th Prague Conference on Information Theory, Statistical Decision Functions, and Random Processes, Prague, 1978, v. B. Akademia, Praha, pp. 267-273.

[135] A. SZULGA, *On minimal metrics in the space of random variables*, Theory Prob. Appl., 27 (1982), pp. 424-430.

[136] A. H. TCHEN, *Inequalities for distributions with given marginals*, Ann. Probab., 8 (1980), pp. 814-827.

[137] V. S. VARADARAJAN, *Weak convergence of measures on separable metric space*, Sankhyā, 19 (1958), pp. 15-22.

[138] R. E. VITALE AND A. C. PIRKIN, *Conditions on the regression function when both variables are uniformly distributed*, Ann. Probab., 4 (1976), pp. 869-873.

[139] JON A. WELLNER, *A Glivenko-Cantelli theorem for empirical measures of independent but nonidentically distributed random variables*, Stochastic Process. Appl., 11 (1981), pp. 309-312.

[140] W. WHITT, *Bivariate distributions with given marginals*, Ann. Statist., 4 (1976), pp. 1280-1289.

[141] V. M. ZOLOTAREV, *General problems of mathematical models*, In: Proceedings of 41st session of ISI, New Delhi, 1977, pp. 382-401.

[142] V. M. ZOLOTAREV, *Ideal metrics in the problems of probability theory*, Austral. J. Statist., 21 (1979), pp. 193-208.

[143] V. L. LEVIN, *The mass transfer problem in topological space and probability measures on the product of two spaces with given marginal measures*, Dokl. Akad. Nauk SSSR, 276 (1984), pp. 1059-1064.

[144] V. L. LEVIN, *The mass transfer problem, strong stochastic domination, and probability measures on the product of two compact spaces with given projections*, Preprint, TsEMI, Moscow. (In Russian.)

[145] D. S. ORNSTEIN, *Ergodic Theory, Randomness and Dynamical Systems*, Yale Math. Monographs No. 5, Yale Univ. Press, New Haven, CT, 1974.

[146] L. GROSS, *Absence of second order phase transition in the Dobrushin uniqueness region*, J. Statist. Phys., 25 (1981), pp. 57-72.

[147] L. GROSS, *Decay of correlation in classical lattice models at high temperature*, Comm. Math. Phys., 68 (1979), pp. 9-27.

[148] D. KLEIN, *Dobrushin uniqueness techniques and decay of correlations in continuum statistical mechanics*, Comm. Math. Phys., 86 (1982), pp. 149-160.

[149] O. E. LANFORD, III, *Entropy and Equilibrium States in Classical Statistical Mechanics*, Lecture Notes in Physics, 20, Springer, Berlin, 1973.

[150] S. L. LEVIN, *Application of Dobrushin's uniqueness theorem to N-vector models*, Comm. Math. Phys., 78 (1980), pp. 65-74.

[151] M. KÜNSCH, *Decay of correlations under Dobrushin's uniqueness condition and its applications*, Comm. Math. Phys., 84 (1982), pp. 207-222.

[152] B. A. SIMON, *A remark on Dobrushin's uniqueness theorem*, Comm. Math. Phys., 68 (1979), pp. 183-185.

[153] R. M. SHORTT, *Combinatorial methods in the study of marginal problems over separable spaces*, J. Math. Anal. Appl., 97 (1983), pp. 462-479.

[154] H. TANAKA, *Probabilistic treatment of Boltzmann equation of Maxwellian molecules*, Z. Wahrsch. Verw. Geb., 46 (1978), pp. 67-105.

[155] V. FRIENDRICH, *Stetige Transportoptiemierung ihre beriehungen zur theorie der höldersteitigen. Funktionen und einige ihrer Anwendungen*, VEB Deutscher Verlag Wissenschaften, Berlin, 1972.