Statistics & Decisions
9, 1991, 327-373

RECENT RESULTS IN THE THEORY OF PROBABILITY METRICS

S. T. Rachev*) and L. Rüschendorf*)

Abstract. The study of metrics in the space of probability measures and random elements has received wide attention and by now there is a wide variety of metrics available for study and use. In this review we discuss the interrelations between metrics of different type, the choice of an appropriate metric for a given approximation problem, the characterization of uniformities and compactness criteria for different metrics as well as applications of the theory of probability metrics to mass transportation problems, characterization of probability distributions and limit theorems for sums and maxima of random elements.

AMS subject classifications: (1984) Primary 60 E 05, 60 B 05,
secondary 60 E 15, 60 B 10

Keywords and phrases: probability metrics, characterization of probability distributions, stability of stochastic models, mass transportation problems, quality usage, Marshall-Olkin distribution, de Finetti's theorem, uniformities

*) Research supported by the Deutsche Forschungsgemeinschaft (DFG)

1. Introduction

The theory of probability metrics (TPM) arises in the same way as any other branch of probability theory. First certain basic properties of the objects to be investigated are found, which are capable of a concrete interpretation, generally in more than one way. The basic properties of the probability metrics *(p - metrics)* come from the fact that they metrize different types of convergences in the space of measures or measurable functions, for example weak convergence, total variation, L_p-convergence, etc. Moreover, due to the probabilistic structure of p- metrics there is a natural correspondence between metrics in the space of probability measures and metrics in the space of random elements; Prohorov distance ↔ Ky Fan distance, Kantorovich- Rubinstein metric ↔ L_1-distance. Concrete interpretation of the basic properties and classification of probability metrics provide the introduction of the TPM. The development of the TPM was closely related to many other branches of probability theory and mathematical statistics such as limit theorems, characterization of probability distributions, robustness, risk theory, queueing theory, quality control and others. p - metrics (i.e. metrics (or semimetrics) in the space of probability measures or in the space of random elements) have been studied by a great number of researchers, among them are Lévy [69], [70], Fréchet [38], Fortét and Mourier [34], Kolmogorov [67], Prohorov [81], Kantorovich and Rubinstein [56], Strassen [114], Dudley [28], [29], and Zolotarev [121], [122], [127].

There are few monographs which treat the p- metrics as a separate subject:

1. Lévy [69] (pp. 199 - 200; definition of Lévy metric);
2. Fréchet [35] (pp. 193 - 195; definition of distances between random variables);
3. Lukacs [71] (Ch. 3; definition and properties of the Ky Fan distance and L_p-distances between random variables);
4. Hennequin and Tortrat [49] (Prohorov distance, total variation distance);
5. Dudley [29], [30] (duality representation of Prohorov and Kantorovich-Rubinstein metrics in the space of probability measures, properties of metrics that metrize the weak convergence);

6. Zolotarev [130] (Ch. 1: introduction to the theory of probability metrics);
7. Kalashnikov and Rachev [55] (Ch. 3, review on the results in the TPM until 1984).

The aim of the present review is to discuss the contemporary state of the TPM mainly paying attention to the wide spectrum of possible applications of the TPM.

For the basic notions of probability metrics we refer to [127], [91], [55] and [93]. Let $\mathfrak{X}(U)$ denote the set of all random variables on a probability space $(\Omega, \mathfrak{U}, P)$ with values in U. We assume that the induced set of distributions of pairs $X, Y \in \mathfrak{X}(U)$ is identical to the set of all distributions on U x U. For a probability metric $\mu(X,Y)$ (cf. (5.1) for definition) the functional

$$\hat{\mu}(P_1, P_2) = \inf \{\mu(X,Y);\ X,Y \in \mathfrak{X}(U),\ P_X = P_1, P_Y = P_2\} \tag{1.1}$$

is called *minimal metric*. In order that $\hat{\mu}$ is a probability metric, some "topological" assumptions on U are to be made (cf. [93]).

2. Metrics Related to Mass Transportation Problems

Probably one of the first metrics in probability theory originated in the works of Monge [74] and Gini [44], [45]. From different aspects they considered in fact the following metric in the distribution function space

$$\varkappa(F,G) = \inf E|X - Y|, \tag{2.1}$$

where the infimum is taken over all joint distributions of pairs (X,Y) with fixed marginal distribution functions F and G.

In 1781, G. Monge proposed in simple prose a seemingly straightforward problem of optimization. It was destined to have wide ramifications. He began his paper on the theory of "clearings and fillings" as follows:

"When one must transport soil from one location to another, the custom is to give the name *clearing* ('deblai') to the volume of the soil that one must transport and the name *filling* ('remblai') to the space that it must occupy after transfer.

The cost of the transportation of one molecule being, all other things equal, proportional to its weight and the interval that it is made to pass-through, and consequently the total cost of transportation having to be proportional to the sum of the projects of the molecules each multiplied by the interval traversed, it follows that the clearing and filling being given in shape and position, it is not indifferent that someone molecule of the clearing be moved to one or another spot of the filling, but rather that there is a certain distribution to be made of the molecules from the first to the second, by which the sum of its projects will be the least possible, and the cost of the total transportation will be a *minimum*." (Monge [74], p. 666)

In 1948 Kantorovich rediscovered the Monge problem (cf. [57]). The abstract form of the *Monge-Kantorovich-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 $\mathfrak{P}(P_1,P_2)$ is the space of all Borel probability measures P on U x U with fixed marginals $P_1(\cdot) = P(\cdot \times U)$ and $P_2(\cdot) = P(U \times \cdot)$. Evaluate the functional

(2.2)
$$\ell_1(P_1,P_2) = \inf\{\int_{U\times U} d(x,y)P(dx,dy):\ P\in\mathfrak{P}(P_1,P_2)\}.$$

The measures P_1 and P_2 may be viewed as the initial and final distribution of mass and $\mathfrak{P}(P_1,P_2)$ as the space of admissible transference plans. If the infimum in (2.2) is realized for some measure $P^* \in \mathfrak{P}(P_1,P_2)$, then P^* is said to be the optimal transference plan. The function d(x,y) can be interpreted as the cost of transferring a unit mass from x to y.

Problem (2.2) was first formulated and studied by Kantorovich for a compact U (cf. [57]). It was shown that

(2.3)
$$\ell_1(P_1,P_2) = B_d(P_1,P_2),$$

where B_d is the Kantorovich metric in the space $\mathfrak{P}_U$ of Borel probability measures on (U,d),

(2.4)
$$B_d(P_1,P_2) := \sup\{|\int_U f d(P_1 - P_2)|:\ f\in \mathrm{Lip}(U)\},$$

and

(2.5)
$$\mathrm{Lip}(U) := \{f: U\to R^1: |f(x) - f(y)| \le d(x,y),\ x,y\in U,\ \sup_{x\in U} |f(x)| < \infty\}.$$

The equality (2.3) for s.m.s. U was proved by Kellerer [59] (see also [89] and [30], Chapter 11, for the case when $\int d(x,a)(P_1 + P_2)(dx) < \infty$).

In many of his writings Corrado Gine (see the survey [45]) discussed the following problem: What is meant b the degree of *concordance* (greater values of X go with greater values of Y)? Gini [44] introduced the concept of *"simple index of dissimilarity"* which coincides with (2.1). Specific contributions to the solution of the Gini problem which is clearly closely related to the Monge-Kantorovich problem (2.2) were made by Hoeffding [50], Fréchet [36], [37], Dall'Aglio [19], [20], [21], Cambanis, Simons and Stout [13], Cambanis and Simons [14], Tchen [115], Rüschendorf [105], [106], [107]. In particular, we refer to the review paper of Rachev [90] on the Monge-Kantorovich and Gini problems. In this section we discuss some recent developments of these problems.

From the *dual* representation (2.3) of the metric ℓ_1 it follows easily that in the special case U = R, d(x,y) = |x - y| we have the following explicit expression for ℓ_1:

(2.6) $$\ell_1(F,G) = \int_{-\infty}^{+\infty} |F(x) - G(x)|dx = \int_0^1 |F^{inv}(t) - G^{inv}(t)|dt,$$

where F^{inv} is the generalized inverse of F. Formula (2.6) shows that the infimum in (2.1) is attained for $X = F^{inv}(V)$, $Y = F^{inv}(V)$, where V is a [0,1]-uniformly distributed random variable. This implies that the optimal association of the "molecules" in the Monge problem can be determined by the "greedy" algorithm, the so-called *northwest corner rule*, that solves transportation problems having a particular structure of the cost and is, moreover, at the heart of many seemingly different problems having an "easy" solution, cf. [51] and [4].

More generally, for any measurable nonnegative cost function c(x,y) on U x U we define the *Kantorovich-functional*

(2.7) $$\hat{\mu}_c(P_1,P_2) = \inf \{ \int_{U\times U} c(x,y)P(dx,dy);\ P \in \mathfrak{P}(P_1,P_2)\},$$

which has the following fundamental representation (cf. [60], [88], [90], [68])

(2.8) $$\hat{\mu}_c(P_1,P_2) = \sup \{\int f\,dP_1 + \int g\,dP_2;\ f \in \mathfrak{L}^1(P_1),\ g \in \mathfrak{L}^1(P_2),\ f(x)+g(y) \le c(x,y),\ \forall\ x,y\}.$$

For special cost functions c(x,y) one can "improve" the representation (2.8) as for example for $c(x,y) = d(x,y)$ in (2.4) (cf. [90], [96]).

Consider a mass transportation problem with cost function

$$c_\varepsilon(x,y) = I\{(x,y) \in U \times U: d(x,y) > \varepsilon\}$$

and masses P_1 and P_2 given on a s.m.s. U. Then for any $\varepsilon > 0$, the following improvement of (2.8), the *Strassen-Dudley-theorem* [114], [28] holds:

$$\hat{\mu}_{c_\varepsilon}(P_1,P_2) = \sup\{(P_1(A) - P_2(A^\varepsilon));\ A \in \mathfrak{B}(U)\}, \tag{2.9}$$

where $A^\varepsilon = \{x: d(x,A) \le \varepsilon\}$ and $\mathfrak{B}(U)$ is the Borel σ-algebra on (U,d). Thus if $K_\lambda(X,Y)$ is the *Ky Fan distance with parameter* $\lambda > 0$ *(distance in probability,* see [33], [82])

$$K_\lambda(X,Y) = \inf\{\varepsilon > 0: Pr(d(X,Y) > \lambda\varepsilon) < \varepsilon\} \tag{2.10}$$

and $\pi_\lambda(P_1,P_2)$ is the Prohorov metric with prameter $\lambda > 0$,

$$\pi_\lambda(P_1,P_2) = \inf\{\varepsilon > 0: \sup_{A\in\mathfrak{B}(U)} (P_1(A) - P_2(A^{\lambda\varepsilon})) < \varepsilon\}, \tag{2.11}$$

then by (2.9)

$$\pi_\lambda(P_1,P_2) = \hat{K}_\lambda(P_1,P_2), \tag{2.12}$$

where

$$\hat{K}_\lambda(P_1,P_2) = \inf\{K_\lambda(X,Y);\ X,Y \in \mathfrak{X}(U),\ P_X = P_1\ P_Y = P_2\}, \tag{2.13}$$

is the minimal metric relative to K_λ.
Letting $\lambda \to 0$ we get

$$\pi_\lambda(P_1,P_2) \to \sigma(P_1,P_2) = \sup_{A\in\mathfrak{B}(U)} |P_1(A) - P_2(A)| \tag{2.14}$$

and thus in the limit one obtains Dobrushin's representation

$$\sigma(P_1,P_2) = \hat{i}(P_1,P_2), \quad \text{where } i(X,Y) = Pr(X \ne Y). \tag{2.15}$$

Letting $\lambda \to \infty$, $\lambda K_\lambda \to L_\infty$ and thus (cf. [83], [28])

$$\hat{L}_\infty(P_1,P_2) = \inf\{\varepsilon > 0: P_1(A) \le P_2(A^\varepsilon), \text{ for all } A \in \mathfrak{B}(U)\}, \tag{2.16}$$

where

$$L_\infty(P) = \text{ess sup}\, d(x,y) = \inf\{\varepsilon > 0;\ P(d(x,y) > \varepsilon) = 0\}. \tag{2.17}$$

Consideration of the cost function $c(x,y) = d^p(x,y)$ for some $0 < p < \infty$ leads to the L_p-*minimal metrics* ℓ_p

$$\ell_p(P_1,P_2) = \inf\{L_p(P): P \in \mathfrak{P}(P_1,P_2)\} = \hat{L}_p(P_1,P_2),\ 0 \le p \le \infty, \tag{2.18}$$

where

$$L_p(P) = [\int_{U\times U} d^p(x,y)P(dx,dy)]^{p'} \quad (0 < p < \infty,\ p' = \min(1,1/p)) \tag{2.19}$$

and $(L_p(P))^{1/p'}$ is the total cost of transportation of P_1 to P_2 by using the transference plan P and cost function $d^p(x,y)$. Letting $p \to 0$ or $p \to \infty$ we get the following limit expressions for L_p:

$$L_o(P) = \int_{U\times U} I\{x \neq y\}P(dx,dy),\ L_\infty(P) = \mathrm{ess}_P \sup d(x,y). \tag{2.20}$$

For $U = R$, $d(x,y) = |x - y|$, $\ell_p = \hat{L}_p$, $1 \le p \le \infty$ admits the explicit representation [13], [83]:

$$\ell_p(P_1,P_2) = \Big[\int_0^1 |F_1^{inv}(t) - F_2^{inv}(t)|^p dt\Big]^{1/p}, \tag{2.21}$$

$$\ell_\infty(P_1,P_2) = \sup_{0\le t\le 1} |F_1^{inv}(t) - F_2^{(inv)}(t)|. \tag{2.22}$$

There was some essential progress on the problem of obtaining explicit expressions and characterizations for the Kantorovich metric in the multidimensional case $U = \mathbb{R}^n$, $n > 1$. A simple completely explicit formula for $\ell_p(P_1,P_2)$ can not be expected, since e.g. the multivariate assignment problem of combinatorial optimization theory is a very special case of this problem. We mention one result on the evaluation of $\ell_2(P,Q)$ where P and Q are distributions in R^n.

For a lower semicontinuous convex function f on R^n let f^* denote the conjugate function

$$f^*(y) = \sup_{x\in R^n} \{\langle x,y\rangle - f(x)\} \tag{2.23}$$

and denote the subdifferential of f in x by

$$\partial f(x) = \{y \in R^n;\ f(z) - f(x) \ge \langle z - x,y\rangle,\ z \in R^n\}, \tag{2.24}$$

(cf. [103]). The elements of $\partial f(x)$ are called subgradients of f at x. Then it holds that for all x,y

$$f(x) + f^*(y) \ge \langle x,y\rangle \tag{2.25}$$

with equality if and only if $y \in \partial f(x)$. The pair (X^*,Y^*) of n-dimensional vectors with marginal distributions P, Q is said to be *optimal* for the Monge-Kantorovich problem with cost function $\|x-y\|_2^2 = \sum_{i=1}^{n} |x_i - y_i|^2$, if (X^*,Y^*) achieves the following infimum in

$$\ell_2^2(P,Q) = \inf \{E\|X-Y\|_2^2 : Pr_X = P, Pr_Y = Q\}. \tag{2.26}$$

The following result is due to Rüschendorf and Rachev [109] extending earlier work of Knott and Smith [66] :

(2.27) (X^*,Y^*) is optimal if and only if $Y^* \in \partial f(X^*)$ (P-a.s.) for some lower semicontinuous convex function f.

The sufficiency of the condition in (2.27) is easy to see. Suppose that X^*, Y^* have distributions P, Q, $Y^* \in \partial f(X^*)$, then for any other rv's X, Y with distributions P, Q we have $E\|X-Y\|^2 = E\|X\|^2 + E\|Y\|^2 - 2E\langle X,Y\rangle$ $\geq E\|X\|^2 + E\|Y\|^2 - 2E(f(X) + f^*(Y)) = E\|X^*\|^2 + E\|Y^*\|^2 - 2E(f(X^*) + f^*(Y^*))$ $= E\|X^* - Y^*\|^2$, since $Y^* \in \partial f(X^*)$.

In particular, if P and Q are Gaussian measures on R^n with means $\vec{m}_1$ and $\vec{m}_2$ and non-singular covariance matrices M_1 and M_2 respectively, one obtains choosing $Y^* = AX^*$, $A = M_1^{1/2}(M_1^{1/2}M_2M_1^{1/2})^{-1/2}M_1^{1/2}$:

$$\ell_2^2(P,Q) = \|\vec{m}_1 - \vec{m}_2\|_2^2 + tr(M_1) + tr(M_2) - 2\,tr\left[\left(\sqrt{M_1}\, M_2 \sqrt{M_1}\right)^{1/2}\right] \tag{2.28}$$

see [77] and [46].

Next let $\mathfrak{Q}(P_1,P_2)$ to be the set of all finite Borel-measures Q on $\mathfrak{B}(U \times U)$ such that

$$Q(A \times U) - Q(U \times A) = P_1(A) - P_2(A),\ A \in \mathfrak{B}(U). \tag{2.29}$$

Define the *Kantorovich-Rubinstein functional*

$$\mathring{\mu}_c(P_1,P_2) = \inf \{\int c(x,y)\, Q(dx,dy);\ Q \in \mathfrak{Q}(P_1,P_2)\}. \tag{2.30}$$

Levin in the early 1970's (see reference [68]) proved the dual representation (analogously to (2.8)):

$$\mathring{\mu}_c(P_1,P_2) = \sup \{\int f\, d(P_1 - P_2);\ f: U \to \mathbb{R}^1,\ f(x) - f(y) \leq c(x,y),\ \forall x,y \in U\} \tag{2.31}$$

for c(x,y) continuous and U compact. Rachev and Shortt [96] got the following strengthened duality representation for symmetric nonnegative cost functions $c(x,y)$ on a separable metric space U satisfying the following conditions:

C.1 $c(x,y) = 0$ iff $x = y$,

C.2 $c(x,y) \le \lambda(x) + \lambda(y)$, $\forall x,y$, for function $\lambda: S \to \mathbb{R}_+$ mapping bounded sets into bounded sets,

C.3 $\sup\{c(x,y);\ x,y \in B_\varepsilon(a),\ d(x,y) \le \delta\} \to 0$ as $\delta \to 0$ for each $a \in U$, $B_\varepsilon(a)$ the ε-ball with center a.

Defining for $f: U \to \mathbb{R}$

(2.32) $$\|f\|_c := \sup\left\{\frac{|f(x)-f(y)|}{c(x,y)};\ x \ne y\right\}$$

the following representation holds:

(2.33) $$\begin{aligned}\overset{\circ}{\mu}_c(P_1,P_2) &= \sup\{|\textstyle\int f\,d(P_1-P_2)|;\ \|f\|_c \le 1\}\\ &= \sup\{\textstyle\int f\,d(P_1-P_2);\ f(x)-f(y) \le c(x,y),\ \forall x,y\},\end{aligned}$$

assuming $\int |x|\,dP_i(x) < \infty$, $i = 1,2$.

While obviously in general

(2.34) $$\overset{\circ}{\mu}_c(P_1,P_2) \le \hat{\mu}_c(P_1,P_2),$$

it follows by comparison with (2.3) that for $c(x,y) = d(x,y)$, $\hat{\mu}_d(P_1,P_2) = \overset{\circ}{\mu}_d(P_1,P_2)$. The cost function

(2.35) $$c_p(x,y) = d(x,y)\max[1, d^{p-1}(x,a), d^{p-1}(y,a)],\ p \ge 1,\ x,y \in \mathbb{R}^1,$$

satisfies C.1 - C.3. From (2.33) we obtain the explicit representation

(2.36) $$\overset{\circ}{\mu}_{c_p}(P_1,P_2) = \int_{-\infty}^{\infty} \max(1, |x-a|^{p-1})\,|F_1(x) - F_2(x)|\,dx,$$

where F_i are the df's of P_i (cf. [89] and for some generalizations [96]). Except for p = 1 an optimal measure Q^* satisfying

(2.37) $$\mu_{c_p}(Q^*) = \int c_p(x,y)\,Q^*(dx,dy) = \overset{\circ}{\mu}_{c_p}(P_1,P_2)$$

is not known. $\overset{\circ}{\mu}_{c_p}$ is for $p \ge 1$ identical to the Fortét-Mourier metric [34]

(2.38) $$FM_p(P_1,P_2) = \sup\{|\textstyle\int_U f\,d(P_1-P_2)|;\ f \in C^p\},$$

where

$$C^p = \{g: U \to \mathbb{R}^1;\ \sup_{r \ge 1} r^{1-p} \sup\{\frac{|g(x)-g(y)|}{d(x,y)},\ x \ne y,\ d(x,a) \le r, d(y,a) \le r\} \le 1\}.$$

A relation between the Kantorovich functional $\hat{\mu}_c$ and the Kantorovich-Rubinstein functional $\mathring{\mu}_c$ can be obtained in the following way. Define for a costfunction $c(x,y) \geq 0$,

$$\tilde{c}(x,y) = \inf\{\sum_{i=1}^{n-1} c(x_i,x_{i+1});\ n \in \mathbb{N},\ x_i \in U,\ x_1 = x,\ x_n = y\}. \tag{2.39}$$

$\tilde{c}(x,y)$ is the minimal cost of a transport from x to y done in several steps. Obviously, $\tilde{c}(x,y) \leq c(x,y)$ and $\tilde{c}$ satisfies the triangle inequality: $\tilde{c}(x,y) \leq \tilde{c}(x,z) + \tilde{c}(z,y)$. If c is symmetric, then $\tilde{c}$ is a (semi-) metric and is obviously the largest (semi-) metric dominated by c. Now suppose that for $\tilde{c}$, c the duality theorem (2.31), (2.33) resp. holds, then we obtain:

$$\begin{aligned} \mathring{\mu}_c(P_1,P_2) &= \sup\{\int f\, d(P_1 - P_2);\ f(x) - f(y) \leq c(x,y),\ \forall x,y\} \\ &= \sup\{\int f\, d(P_1 - P_2);\ f(x) - f(y) \leq \tilde{c}(x,y),\ \forall x,y\} \\ &= \hat{\mu}_{\tilde{c}}(P_1,P_2). \end{aligned} \tag{2.40}$$

From (2.40) for c symmetric, we obtain

$$\begin{aligned} \mathring{\mu}_c(P_1,P_2) &= \sup\{|\int f\, d(P_1 - P_2)|;\ |f(x) - f(y)| \leq \tilde{c}(x,y),\ \forall x,y\} \\ &= \hat{\mu}_{\tilde{c}}(P_1,P_2). \end{aligned} \tag{2.41}$$

If Q^* is an optimal measure w.r.t. $\tilde{c}$ we obtain that $c(x,y) = \tilde{c}(x,y)$ a.s. w.r.t. Q^*.This gives a natural explanation of the relevance of $\mathring{\mu}_c$ for transportation problems. A somewhat different interpretation of $\mathring{\mu}_c$ can be found in Kemperman [61] (multistage shipping). In linear programming the discrete analogon is known as network flow problem. Kantorovich and Rubinstein [56] studied in 1957 a modification with exactly n-stages of the transportation. In terms of rv's we may also give the following representation.

$$\begin{aligned} \mathring{\mu}_c(P_1,P_2) &= \hat{\mu}_{\tilde{c}}(P_1,P_2) = \inf\{E\tilde{c}(X_1,X_2);\ P_{X_1} = P_1,\ P_{X_2} = P_2\} \\ &= \inf\{E[c(X_1,X_2) + c(X_2,X_3) + \dots + c(X_{n-1},X_n)];\ n \in \mathbb{N} \\ &\qquad P_{X_1} = P_1,\ P_{X_n} = P_2,\ X_i \text{ any rv's},\ 2 \leq i \leq n-1\} \\ &= \inf\{\int c(x,y)Q(dx,dy);\ Q \in \mathfrak{Q}(P_1,P_2),\ \mathfrak{Q}(U \times U) \in \mathbb{N}\}, \end{aligned} \tag{2.42}$$

here $Q = \sum_{i=1}^{n-1} P_{X_i,X_{i+1}}$. From (2.41) obviously $\mathring{\mu}_c$ is a semimetric on $\mathfrak{P}(U)$ if c is symmetric (cf. also [28], Lemma 20.2, [86]). If $c(x,y) = d^p(x,y)$, $p > 1$, $U = \mathbb{R}^k$, then $\tilde{c}(x,y) = 0$ and, therefore, by (2.41) $\mathring{\mu}_c(P_1,P_2) = 0$. This shows a striking difference between $\hat{\mu}_c$ and $\mathring{\mu}_c$.

Inequalities between ℓ_p, $\overset{\circ}{\mu}_{c_p}$ and other metrics on $\mathfrak{P}(U)$ are studied in [83], [84 , [85] . In particular for any $P_o \in \mathfrak{P}(U)$, $\mathfrak{D}(P_o, \ell_p) = \{P \in \mathfrak{P}(U); \ell_p(P, P_o) < \infty\}$ is ℓ_p-complete and for $P_o = \delta_a$ the following convergence criterion holds on $\mathfrak{D}(\delta_a, \ell_p)$:

(2.43)
$$\ell_p(P_n, P) \to 0 \Leftrightarrow \overset{\circ}{\mu}_{c_p}(P_n, P) \to 0$$
$$\Leftrightarrow P_n \to P \text{ (weakly) and } \int d^p(x,y)(P_n - P)(dx) \to 0.$$

For $P_o \neq \delta_a$ a corresponding characterization is unknown. A compactness criterion for ℓ_∞ is unknown in the general case, for $U = \mathbb{R}^1$ cf. [101].

We finally describe some applications of the Kantorovich resp. Kantorovich-Rubinstein functionals $\hat{\mu}_c$, $\overset{\circ}{\mu}_c$.

2.1. Classification Problem, Asignment Problem

Suppose that n individuals should be classified or asigned to n jobs. After a series of tests one knows the empirical measure $P_1(A)$, for $A \subset U$, the set of "qualities" which are of interest for the jobs. $P_1(A)$ denoting the relative number of individuals with qualities in the set A. $P_2(A)$ represents the desired distribution of the qualities for the jobs. A distance $d(x,y)$, $x,y \in U$, is measuring the ability for an individual with quality x to cope with the needed quality y of a job (e.g. $U = \mathbb{R}^k$, $d(x,y) = \|x - y\|$). Each element $P \in \mathfrak{P}(P_1, P_2)$ describes a classification. $P(A \times B)$ is the number of individuals with qualities in A which are assigned to jobs with qualities in B. Then $\hat{\mu}_c(P_1, P_2)$ is the optimal amount of disclassification. In a model where individuals can gradually change their qualities to fit to the different jobs in a sequence of n retraining stages, a classification is given by an element $Q \in \mathfrak{Q}(P_1, P_2)$ and the optimal disclassification amount is given by $\overset{\circ}{\mu}_c(P_1, P_2) \leq \hat{\mu}_c(P_1, P_2)$ (cf. [95]).

2.2. Optimal Allocation Policy

Karatzas [58] (see also the general discussion in Whittle [119], pp. 210 - 211) considers d "medical treatments" (or "projects" or "investigations") with the state of the j-th of them (at time $t \geq 0$) denoted by $x_j(t)$. At each instant of time t, we are allowed to use only one medical treatment denoted by i(t), which then evolves according to some Markovian rule; meanwhile, the states of all other projects remain frozen. If i(t) = j, one acquires an instant reward equal to $h(j,x_j(t))$ per unit time, discounted by the factor $e^{-\alpha t}$. The stochastic control problem is then to choose the sequential "allocation policy" {i(t), t ≥ 0} in such a way as to maximize the expected discounted reward $E \int_0^\infty e^{-\alpha t} h(i(t),x_{i(t)}(t))dt$.

We will now consider the situation, when we are allowed to use a combination of different medical treatments (say for brevity, medicines) denoted by $M_1,\dots,M_d$. Let d = 2 and $U = U_1 \times \mathbb{R}_+$, U_1 the space of relevant parameters of the patient, $\mathbb{R}_+$ the time. Let $P_i(A \times B)$ be the total quantity of medicine M_i which should be given to a patient with parameters in A in time B. P_i can be normalized by 1. An admissible allocation (treatment) policy describes the combination of medicines over the time and is given by an element $P \in \mathfrak{P}(P_1,P_2)$ (resp. $Q \in \mathfrak{Q}(P_1,P_2)$). If we can specify a reward function $c((x_1,t_1), (x_2,t_2))$ **describing the interaction of the medicines, then we can** formulate the question of an optimal policy of combination of medicines (depending on time and patient's status) as a Monge-Kantorovich problem with a treatment policy $P \in \mathfrak{P}(P_1,P_2)$ or in the case of multistaged treatment as a Kantorovich-Rubinstein problem with treatment policy $Q \in \mathfrak{Q}(P_1,P_2)$.

2.3. Generalized Kantorovich Theorem and Optimal Quality Usage

It is common practice to describe the quality of an item of a product by a collection of its characteristics $x = (x_1,\dots,x_m) \in \mathbb{R}^m$. The quality of all produced items of a given type is described by a probability measure μ(A), $A \in \mathfrak{B}^m$ the algebra of Borel measurable sets in $\mathbb{R}^m$. The measure μ(A) represents the proportion of items with quality x ∈ A. On the other hand the usage (consumption) of items can be represented by another probability measure ν(B), $B \in \mathfrak{B}^m$, where ν(B) describes the necessary consumption product with quality characteristics x ∈ B. We call μ(A) the *production measure* and ν(B) the *consumption measure*, $A,B \in \mathfrak{B}$ and assume that $\mu(\mathbb{R}^m) = \nu(\mathbb{R}^m) = 1$, [57]. Let c(x,y) denote the cost (or degree of unpleasantness) to satisfy a demand y with a produced quality x possibly different from y.

In most practical cases the information about production and consumption quality concerns only the marginal features of the production and consumption measures. We denote the i^{th} marginal measure of production quality by $\mu_i(A_i)$ and the j^{th} marginal measure of the consumption quality by $\nu_j(B_j)$, i.e.

(2.44)
$$\mu_i(A_i) = \mu(\mathbb{R}^{i-1} \times A_i \times \mathbb{R}^{m-i}),\ A_i \in \mathfrak{B}^1$$
$$\nu_j(B_j) = \nu(\mathbb{R}^{j-1} \times B_j \times \mathbb{R}^{m-j}),\ B_j \in \mathfrak{B}^1$$

[27]. We say that a probability measure P on $\mathfrak{B}^{2m}$ is a *weakly admissible plan* when it satisfies the marginal conditions

(2.45)
$$P(\mathbb{R}^{i-1} \times A_i \times \mathbb{R}^{2m-i}) = \mu_i(A_i),\ i = 1,\dots,m;$$
$$P(\mathbb{R}^{m+j-1} \times B_j \times \mathbb{R}^{m-j}) = \nu_j(B_j),\ j = 1,\dots,m.$$

Denote by $\overline{\mathfrak{P}}(\mu_1,\dots,\mu_m;\nu_1,\dots,\nu_m)$ the collection of all weakly admissible plans, then

(2.46)
$$\mathfrak{P}(\mu,\nu) \subseteq \overline{\mathfrak{P}}(\mu_1,\dots,\mu_m;\nu_1,\dots,\nu_m).$$

A weakly admissible plan P^o is called *optimal* w.r.t. a cost function c if it satisfies the relation

(2.47)
$$\mu_c(P^o) = \min_{P\in\overline{\mathfrak{P}}} \mu_c(P).$$

(2.47) implies that

(2.48)
$$\mu_c(P^o) \le \mu_c(P^*) := \hat{\mu}_c(\mu,\nu).$$

Equality holds in (2.48) e.g. if $\mu = \otimes \mu_i$, $\nu = \otimes \nu_i$ and $c(x,y) = \sum_{i=1}^m c_i(x_i,y_i)$. For the determination of $\mu_c(P^o)$ there are several explicit results mainly based on the generalized Kantorovich theorem

(2.49)
$$\mu_c(P^o) = \sup\{\sum_{i=1}^m \int f_i d\mu_i + \sum_{j=1}^m \int g_j d\mu_j;\ f_i \in L^1(\mu_i),\ g_i \in L^1(\mu_i) \sum_{i=1}^m (f_i(x_i) + g_i(y_i)) \le c(x,y),\ x,y \in \mathbb{R}^m\},$$

holding for general costfunctions c (as in (2.8)) (cf. [39], [106], [107], [60], [89]). In particular it was proved in Rüschendorf [106] that for any $A_i, B_i \in \mathfrak{B}(U)$ we have the sharpness of the Fréchet-bounds

(2.50) $\sup\{P(A_1 \times \dots \times A_m \times B_1 \times \dots \times B_m);\ P \in \overline{\mathfrak{P}}\} = \min\{\mu_i(A_i), \nu_i(B_i), 1 \le i \le m\}$,

(2.51) $\inf\{P(A_1 \times \dots A_m \times B_1 \times \dots \times B_m);\ P \in \overline{\mathfrak{P}}\} = \left(\sum_{i=1}^{m} (\mu_i(A_i) + \nu_i(B_i)) - (2m-1)\right)_+$.

For a recent review of these and related Fréchet-type bounds we refer to [111].

3. Uniformities for Weak Convergence and Convergence in Probability

There have been some interesting recent results on the characterization of the uniform structure of probability metrics (cf. [30], Chapter 11, [25], [17], [90], [100], [101]). In contrast to the characterization of convergent sequences and to the description of compactness criteria (as e.g. in (2.43)) we will consider criteria for convergence in the noncompact (merging) case. If $\mu(X,Y)$ and $\nu(X,Y)$ are probability metrics, then we shall consider criteria for convergence of $\mu(X_n,Y_n) \to 0$, respectively, the question of uniform comparability: $\mu(X_n,Y_n) \to 0$ implies $\nu(X_n,Y_n) \to 0$.

Consider for $\lambda \in (0,\infty)$ the Prohorov type metric π_λ from (2.11); $\pi_1 = \pi$ is the usual Prohorov metric. It is well known that for separable metric spaces π, π_λ metrize the topology of weak convergence on $\mathfrak{P}(U)$, i.e. if μ_n, $\mu \in M^1(S)$, $n \in \mathbb{N}$, then:

(3.1) $$\mu_n \xrightarrow{\mathfrak{D}} \mu \text{ (weak convergence)} \Leftrightarrow \pi_\lambda(\mu_n,\mu) \to 0.$$

Moreover the following sequence of equivalences holds (cf. Dudley [30], p. 310):

a) $\mu_n \xrightarrow{\mathfrak{D}} \mu$,

b) $\int f\,d\mu_n \to \int f\,d\mu$, $\forall f \in BL(U,d)$ the class of bounded Lipschitz-functions with $\|f\|_{BL} = \|f\|_L + \|f\|_\infty < \infty$,

(3.2) c) $\beta(\mu_n,\mu) := \sup\{\int f\,d(\mu_n - \mu);\ \|f\|_{BL} \le 1\} \to 0$,

d) $\pi(\mu_n,\mu) \to 0$,

e) there exists U-valued random variables $X_n \overset{d}{\sim} \mu_n$, $X \overset{d}{\sim} \mu$ on some probability space $(\Omega,\mathfrak{A},P)$ such that $d(X_n,X) \to 0$ a.s.

Some interesting applications of (3.2) to Glivenko-Cantelli theorems, functional central limit theorems and stability of queueing systems are described in [29] and [90]. The equivalence of a), d), e) is the famous Skorohod, Strassen, Dudley, Wichura a.s. convergence theorem. The following "extension" of this equivalence can be proved in the "noncompact" case (cf. [101]).

<u>THEOREM</u> 3.1. For any $\lambda \in (0,\infty)$ and $\mu_n, \nu_n \in \mathfrak{P}(U)$ holds:

(3.3) $\pi_\lambda(\mu_n,\nu_n) \to 0$, if and only if there exist U-valued rv's $X_n \overset{d}{\sim} \mu_n$, $Y_n \overset{d}{\sim} \nu_n$ on a probability space $(\Omega,\mathfrak{U},P)$ such that: $d(X_n,Y_n) \to 0$ [P]. □

For $\pi = \pi_1$, (3.3) was proved independently by Rachev, Rüschendorf and Schief [100] and Dudley [30]. As the a.s. convergence theorem (3.2),e) in the "compact" case, (3.3) has interesting applications in statistics e.g. to establish general versions of the δ-method (the differentiation approach to convergence theorems).

For the limiting cases of the Prohorov-type metric π_λ as $\lambda \to 0$ or $\lambda \to \infty$, $\pi_0 = \sigma$ (cf. (2.14)), $\pi_\infty = \hat{L}_\infty$ (cf. (2.16)) we have the following analog results (cf. [101]):

(3.4) $\sigma(\mu_n,\nu_n) \to 0$ iff for some versions $X_n \overset{d}{\sim} \mu_n$, $Y_n \overset{d}{\sim} \nu_n$ $I\{X_n \neq Y_n\} \to 0$ a.s.

and

(3.5) $\hat{L}_\infty(\mu_n,\mu_n) \to 0$ iff for some versions $X_n \overset{d}{\sim} \mu_n$, $Y_n \overset{d}{\sim} \nu_n$ ess sup $d(X_n,Y_n) \to 0$.

To characterize the uniformity of the Prohorov metric π we assume w.l.g. that the metric d is bounded (otherwise define e.g. $d^*(x,y) = \min(1,d(x,y))$), then we have the following equivalence analogously to (3.2) (cf. [101]):

<u>THEOREM</u> 3.2.

a) $\pi(\mu_n,\nu_n) \to 0$,

	b)	$\beta(\mu_n,\nu_n) \to 0,$
(3.6)	c)	$d_L(\mu_n,\nu_n) = \sup\{\lvert\int f\,d(\mu_n - \nu_n)\rvert;\ \|f\|_L \le 1\} \to 0,$
	d)	$\widehat{L}_1(\mu_n,\nu_n) \to 0,$
	e)	$\widehat{L}_p(\mu_n,\nu_n) \to 0,\quad \forall\, 0 < p < \infty.$ □

The equivalence of uniformities w.r.t. two probability metrics σ_1, σ_2 is a consequence of inequalities of the form:

$$\psi(\sigma_1(\mu,\nu)) \le \sigma_2(\mu,\nu) \le \varphi(\sigma_1(\mu,\nu)) \tag{3.7}$$

for some functions ψ, φ continuous in 0, $\varphi(0) = \psi(0) = 0$, $\varphi(x) \neq 0$, $\psi(x) \neq 0$. For several inequalities of this type cf. [101].

• For the convergence in probability metrized by the Ky-Fan distance $K = K_1$ (cf. (2.10)) we have some results analogously to (3.4), (3.6).

THEOREM 3.3. [101] Let X_n, Y_n be U-valued rv's.

1. If U is compact, then the following are equivalent:

	a)	$E\lvert f(X_n) - f(Y_n)\rvert \to 0,\ \forall f \in BL(U,d)$
(3.8)	b)	$d_{BL}(X_n,Y_n) := \sup\{E\lvert f(X_n) - f(Y_n)\rvert;\ f \in BL(U,d)\} \to 0$
	c)	$K(X_n,Y_n) \to 0$. □

2. (3.8) also holds true for general U if P^{Y_n} has densities h_n w.r.t. a dominating measure μ and $|h_n| \le h$ for some integrable h.
3. $K(X_n,Y_n) \to 0$ is equivalent to the existence of rv's X_n, Y_n, such that (X_n,Y_n) have the same distribution as (X_n,Y_n) and $d(\tilde{X}_n,\tilde{Y}_n) \to 0$ a.s.

4. "Appropriate" Metrics for Approximation of Distributions and Stability of Characterizations of Exponential, Marshall-Olkin and Beta Distributions

Any concret stochastic approximation problem requires an "appropriate" or "natural" metric (topology, convergence, uniformities, etc.) having properties which are helpful in solving the problem. If one needs to develop the solution of the approximation problem in terms of another metric (topology, etc.) the transition is carried out by using general relationships between metrics (topologies, etc.). This two stage approach (selection of the appropriate metric and comparison of metrics) is the basis of the TPM [122], [127], [129].

Since there is not a satisfactory definition of a "natural" metric, we shall use different examples to explain this approach. The first two examples deal with approximation of exponential families of distributions.

EXAMPLE 4.1. *Robustness of χ^2-test of exponentiality*
Suppose that Y is exponentially distributed with density (p.d.f.) $f_Y(x) = 1/a\, e^{-x/a}$, $(x \geq 0;\ a > 0)$. To perform hypothesis tests on a, one makes use of the fact that, if $Y_1, Y_2, \ldots, Y_n$ are n independent, identically distributed random variables, each with p.d.f. f_Y, then $2 \sum_{i=1}^{n} Y_i / a \approx \chi^2_{2n}$. In practice, the assumption of exponentiality is only an approximation; it is therefore of interest to enquire how well the χ^2_{2n} distribution approximates that of $2 \sum_{i=1}^{n} X_i / a$, where $X_1, X_2, \ldots, X_n$ are n independent, identically distributed, non-negative random variables with common mean a, representing a "perturbation", in some sense, of exponential random variables with the same mean. The usual approach requires one to either make an assumption concerning the class of random variables representing the possible "perturbations" of the exponential distribution or to identify the nature of the "mechanism" causing the "perturbation".

The case when X's belong to an aging class distributions. A nonnegative random variable X with distribution function F is said to be HNBUE *(harmonic new better than used in expectation)* if $\int_x^{\infty} \overline{F}(u)du \leq a\, e^{-x/a}$ for all $x \geq 0$, where $a = E(X)$ and $\overline{F} = 1 - F$. It is easily seen that if X is HNBUE, moments of all orders exist. Similarly, X is said to be HNWUE *(harmonic new worse than used in expectation)* if $\int_x^{\infty} \overline{F}(u)du \geq a\, e^{-x/a}$ for all $x \geq 0$ assuming that a is finite. See [55] and [65] for further details of HNBUE and HNWUE distributions. The class of HNBUE (HNWUE) distributions include all the standard "ageing" ("anti-ageing") classes, IFR, IFRA, NBU and NBUE (DFR, DFRA, NWU and NWUE).

It is well known that *if X is HNBUE with $a = EX$ and $\sigma^2 = var\ X$ then X is exponentially distributed if and only if $a = \sigma$.* To investigate stability of this characterization we must select a metric $\mu(X,Y) = \mu(F_X, F_Y)$ in the distribution functions space $\mathfrak{F}(\mathbb{R})$ such that

a) μ guarantees the convergence in distribution plus convergence of the first two moments;

b) one can construct inequalities

(4.1) $\varphi_1(|a-\sigma|) \le \mu(X,E(a)) \le \varphi_2(|a-\sigma|)$, ($X \in$ HNBUE, $EX = a$, $\operatorname{var} X = \sigma^2$,

where φ_i are continuous increasing functions with $\varphi(0) = 0$, $E(a)$ denotes an exponential variate with mean a.

Clearly, the most appropriate metric μ should satisfy a) and b) with $\varphi_1 \equiv \varphi_2$. Such a metric is

(4.2) $$\zeta_2(X,Y) = \sup_{f \in F_2} |E(f(X) - f(Y))| = \int_{-\infty}^{+\infty} \Big| \int_{-\infty}^{x} (F_X(t) - F_Y(t))dt \Big| dx,$$

where F_2 is the class of all functions f having almost everywhere second derivative f" and $|f''| \le 1$ a.e.; for details, see [76], and [5]. The metric ζ_2 is called Zolotarev ζ_2-metric (see [126]). From the first representation of ζ_2 it follows that ζ_2-convergence implies convergence of the second moments. Moreover, if L is the Levy metric on $\mathfrak{F}(\mathbb{R})$ (see [71], Sec. 3.4)

(4.3) $$L(X,Y) = \inf\{\varepsilon > 0;\ F_X(x-\varepsilon) - \varepsilon \le F_Y(x) \le F_X(x+\varepsilon) + \varepsilon \quad \text{for all } x \in \mathbb{R}\},$$

then

(4.4) $$L(X,Y) \le [4\zeta_2(X,Y)]^{1/3},$$

see [47], [72]. Thus ζ_2-convergence preserves the converge in distribution, and so a) holds. Concerning b), we use the second representation of ζ_2 to get

(4.5) $$\zeta_2(X,Y) = \int_0^{\infty} \Big| \int_x^{\infty} \overline{F}_X(t)dt - a e^{-x/a} \Big| dx = \int_0^{\infty} [a e^{-x/2} - \int_x^{\infty} \overline{F}_X(t)dt]dx$$
$$= \frac{1}{2}(a^2 - \sigma^2); \text{ for X is HNBUE, } Y := E(a).$$

Now if one studies the stability the above characterization in terms of a "traditional" metric as the uniform one

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

then one simply compares ζ_2 with ρ. Namely, by the well-known inequality between the Levy distance L and the Kolmogorov distance ρ, we have

(4.7) $$\rho(X,Y) \le [1 + \sup_x f_X(x)] L(X,Y)$$

if $f_X = F'_X$ exists. Thus, by (4.4) and (4.7)

$$\rho(X,Y) \le [1 + \sup_t f_{cX}(t)][4\zeta_2(cX,cY)]^{1/3} = (c^{2/3} + M_X c^{-1/3})[4\zeta_2(X,Y)]^{1/3}$$
for any $c > 0$,

where $M_X = \sup_t f_X(t)$. Minimizing the right hand side of the last inequality with respect to c, we obtain

$$\rho(X,Y) \le 3M_X^{2/3}(\zeta_2(X,Y))^{1/3}. \tag{4.8}$$

Thus, for any $X \in$ HNBUE with $EX = a$, $\operatorname{var} Y = \sigma^2$

$$\rho(X,E(a)) \le 3(\alpha/2)^{1/3}, \ \alpha = 1 - \sigma^2/a^2. \tag{4.9}$$

Note that the order 1/3 of α is sharp, see [18].

Next using the "natural" metric ζ_2, we derive a bound on the uniform distance between the χ^2_{2n} distribution and the distribution of $2\sum_{i=1}^n X_i/a$, assuming that X is HNBUE. Define $\overline{X}_i = (X_i - a)/a$ and $\overline{Y}_i = (Y_i - a)/a$ $(i = 1,2,\dots,n)$ and write $W_n = 2\sum_{i=1}^n X_i/a$, $\overline{W}_n = \sum_{i=1}^n \overline{X}_i/\sqrt{n}$, $Z_n = 2\sum_{i=1}^n Y_i/a$ and $\overline{Z}_n = \sum_{i=1}^n \overline{Y}_i/\sqrt{n}$. Let $f_{\overline{Z}_n}$ denote the p.d.f. of $\overline{Z}_n$ and let $M_n = \sup_x f_{\overline{Z}_n}(x)$. Then by (4.8),

$$\rho(\overline{W}_n,\overline{Z}_n) \le 3M_n^{2/3}[\zeta_2(\overline{W}_n,\overline{Z}_n)]^{1/3}.$$

Now we use the fact that ζ_2 is *"ideal metric of order 2"* [122], i.e. *for any vectors* $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ *with independent components and constants* $c_1,\dots,c_n$,

$$\zeta_2\Big(\sum_{i=1}^n c_iX_i, \sum_{i=1}^n c_iY_i\Big) \le \sum_{i=1}^n |c_i|^2\zeta_2(X_i,Y_i). \tag{4.10}$$

Thus $\zeta_2(\overline{W}_n,\overline{Z}_n) \le \zeta_2(X,Y)/a^2$, and finally the required estimate is

<u>THEOREM</u> 4.1. [5]

$$\rho(W_n,Z_n) \le \frac{3}{2^{1/3}}(1-(\sigma/a)^2)^{1/3}, \tag{4.11}$$

where

$$M_n < \sqrt{\frac{n}{2\pi(n-1)}}\, e^{-1/(12n-11)}. \qquad \square$$

In the same way one can treat the case $X \in$ HNWUE [5]. If one makes no assumptios concerning X, it is necessary to make an assumption concerning the "mechanism" by which the exponential distribution is "perturbed". Then using the estimates for ζ_2 we can deduce bounds for $\rho(W_n, Z_n)$ in the case of the three most common possible "mechanism": contamination by mixture, contamination by an additive error and right-censoring [5].

<u>EXAMPLE</u> 4.2. *Stability of a Characterization of the Marshall-Olkin Distribution*

The derivation of the estimates (4.9), (4.11) is just a simple example of how one can use the TPM. While in the case of (4.9), one can get similar results by the traditional methods (see Daley [18]) in order to study multivariate versions of (4.9) one cannot use the standard technique [11], [12], [18].

Further the TPM is utilized to analyze the stability of a new characterization of the bivariate Marshall-Olkin distribution: We show that if a distribution possesses a certain bivariate NBU property, it is Marshall-Olkin if and only if a given function of the first and second moments and the hazard rates at the origin vanishes.

Recall that if $\tilde{G}(x,y) = P\{X_1 > x, X_2 > y\}$ denotes the bivariate survivor function of (X_1,X_2), then the bivariate Marshall-Olkin distribution (BMOD) is defined by

(4.12) $$\tilde{G}(x,y) := \exp\{-\lambda_1 x - \lambda_2 y - \lambda_{12} \max(x,y)\}$$

for $\lambda_1, \lambda_2 > 0$, $\lambda_{12} \geq 0$, [73].

Let **B** denote the class of all bivariate survivor functions of pairs of non-negative random variables. For $G \in$ **B**, define the hazard vector $(h_1(t), h_2(t)) = \nabla[-\log G(t,t)]$, assuming that this exists, and write $H_1(x,y) = -\partial \log G(x,y)/\partial y$ and $c = H_1(0,0) + H_2(0,0)$. Fixing $h_1(0)$, $h_2(0)$ and c

(4.13) $$\tilde{G}(x,y) = \begin{cases} \exp\{-cy-h_1(0)(x-y)\} & \text{if } x \geq y \\ \exp\{-cx-h_2(0)(y-x)\} & \text{if } x \leq y \end{cases},$$

i.e. $\tilde{G}$ is the bivariate survivor function of the Marshall-Olkin distribution with $\lambda_1 = c - h_2(0)$, $\lambda_2 = c - h_1(0)$ and $\lambda_{12} = h_1(0) + h_2(0) - c$. Notice that if (X_1,X_2) has survivor function $\tilde{G}$, then $E(X_1) = 1/h_1(0)$ and $E(X_1^2) =$

$2/[h_i(0)]^2$ $(i = 1,2)$ and $E(X_1X_2) = [1/h_1(0) + 1/h_2(0)]/c$, where $(h_1(0), h_2(0)) = \nabla[-\log G(0,0)]$.

Suppose that $G \in \mathbf{B}$, *the survival function of* (X_1, X_2), *satisfies the inequalities*

(i) $G(x+t, y+t) \le G(x,y)G(t,t)$ for all $x,y,t \ge 0$,

(ii) $G_i(x+t) \le G_i(x)G_i(t)$ for all $x,t \ge 0$, where G is the survivor function of X_i $(i = 1,2)$.

Then G *is said to be weakly bivariate NBU (WBNB).*

Obviously any bivariate distribution with increasing failure rate (BIFR) [3] is WBNBU. However, the bivariate Weibull-Marshall-Olkin distribution [73] $G_\alpha(x_1,x_2) = \exp\{-\lambda_1 x_1^\alpha - \lambda_2 x_2^\alpha - \lambda_{12}\max(x_1^\alpha, x_2^\alpha)\}$, $x_i \ge 0$, $\alpha > 0$ is WBNBU but not BIFR.

For any survival function G in $\mathbf{B}$ define

$$\begin{aligned}\beta(G) := \{&[1/h_1(0) - E(X_1)] + [1/h_2(0) - E(X_2)] \\ &+ [2E(X_1) - h_1(0)E(X_1^2)] + [2E(X_2) - h_2(0)E(X_2^2)] \\ &+ 4[E(X_1) + E(X_2) - cE(X_1X_2)]\}/c.\end{aligned} \tag{4.14}$$

Clearly, $\beta(G) = 0$ if G is BMOD. In general, $\beta(G)$ represents a measure of closeness of the moments of G to the corresponding moments of the BMOD $\tilde{G}$ defined by (4.13).

<u>THEOREM</u> 4.2. [6]. If G is WBNBU, then G is Marshall-Olkin if and only if $\beta(G) = 0$. Moreover, the uniform distance between $G \in$ WBNBU and $\tilde{G}$ can be estimated by $\beta(G)$ as follows:

$$\rho(G, \tilde{G}) \le (1 + c^2 e^c)[\beta(G)]^{1/3}. \tag{4.15}$$

The proof of (4.15) follows the same two stage approach as in Example 4.1. Here, the "natural" metric in terms of which an inequality is easy to prove is

$$\zeta_2^*(G_1, G_2) = \int_0^\infty \int_0^\infty |G_1(x,y) - G_2(x,y)|\,dx\,dy. \tag{4.16}$$

(4.15) follows from the following three inequalities:

a) (4.17) $\quad \zeta_2^*(G,\tilde{G}) \le \beta(G)$,

b) (4.18) $\quad \zeta_2^*(G_1,G_2) \ge (L(G_1,G_2))^3, \ \forall\, G_1,G_2$,

where

$$L(G_1,G_2) = \inf\{\varepsilon > 0;\, G_1(x-\varepsilon,y-\varepsilon) - \varepsilon \le G_2(x,y) \le G_2(x+\varepsilon,y+\varepsilon) + \varepsilon,\ \forall\, x,y\} \tag{4.19}$$

is the Levy-metric.

c) (4.20) $\quad \rho(G,\tilde{G}) \le (1 + c^2 e^c) L(G,\tilde{G})$.

<u>EXAMPLE</u> 4.3. *Stability of Characterizations of Beta-distributions and Stability of de Finetti's Theorem*

Consider the following question: *Let $\zeta_1,\zeta_2,\ldots$ be a sequence of independent identically distributed (i.i.d.) positive random variables (r.v.'s) with d.f. F satisfying the normalization $E\zeta_1^p = 1$, $\infty > p > 1$ and define*

$$X_{k,n,p} = \sum_{j=1}^{k} \zeta_j^p \,/ \sum_{j=1}^{n} \zeta_j^p, \quad 1 \le k \le n,\ n \in \mathbb{N} := \{1,2,\ldots\}. \tag{4.21}$$

Does there exist a (unique?) d.f. $F = F_p$ such that $X_{k,n,p}$ has a Beta $B(\frac{k}{p},\frac{n-k}{p})$- distribution for any $k \le n$, $n \in \mathbb{N}$? It is well-known that F_1 is the standard exponential distribution and F_2 is the absolute value of a standard normal r.v. (see for example [16], Sec. 18, [26]). As we will see the affirmative answer of this question and its stability with respect to a small departures from the Beta-distribution will lead to satisfy theorems for de Finetti type characterizations of scale mixtures and exponential type families. For references on de Finetti's theorem we recommend Diaconis and Freedman [24], [26]. The results included in this section are due to Rachev and Rüschendorf [94].

<u>THEOREM</u> 4.3. For any $0 < p < \infty$ there exists exactly one distribution $F = F_p$, such that for all $k \le n$, $n \in \mathbb{N}$, $X_{k,n,p}$ has a $B(\frac{k}{p}, \frac{n-k}{p})$-distribution. F_p has the density

$$f_p(x) = \frac{p^{1-1/p}}{\Gamma(1/p)} \exp\left(-\frac{x^p}{p}\right), \quad x \ge 0. \qquad \square$$

To get a meaningful result for $p=\infty$, let β be a $B(\frac{k}{p},\frac{n-k}{p})$-distributed r.v. and define $\gamma_{k,n,p}=\beta^{1/p}$, then $\gamma_{k,n,p}$ has a density given by

$$f_{\gamma_{k,n,p}}(x)=B(\tfrac{k}{p},\tfrac{n-k}{p})\,p\,x^{k-1}(1-x^p)^{(n-1)/p},\ 0\le x\le 1.$$

Let $\gamma_{k,n,\infty}$ be the weak limit of $\gamma_{k,n,p}$ as $p\to\infty$, i.e.

$$P(\gamma_{k,n,\infty}\le x)=\begin{cases}\frac{n-k}{n} & \text{if } 0\le x>1\\ 1 & \text{if } x\ge 1\end{cases}. \tag{4.22}$$

<u>THEOREM</u> 4.4. Let $\zeta_1,\zeta_2,\dots$ be a sequence of positive iid r.v.'s and let $X_{k,n,\infty}=\bigvee_{i=1}^{k}\zeta_i/\bigvee_{i=1}^{n}\zeta_i$ ($\vee\zeta_i:=\max\zeta_i$). Then $X_{k,n,\infty}$ and $\gamma_{k,n,\infty}$ are equally distributed for any $k\le n$, $n\in\mathbb{N}$, if and only if ζ_1 is uniformly distribution on $[0,1]$.

□

By Theorem 4.4 one can say that in the limit case $p=\infty$ the required F_∞ is the uniform $[0,1]$-distribution.

Let us now look at the stability of the characterization of F_p, $p\in(0,\infty]$.

Once again we follow the two stage approach as in Example 4.1. We consider a sequence $\tilde\zeta_1,\tilde\zeta_2,\dots$ of i.i.d. nonnegative r.v.'s with common df $\tilde F_p$ close to F_p in the sense that the uniform distance $\rho:=\rho(\tilde\zeta_1,\zeta_1)=\rho(\tilde F_p,F_p)$ (see (4.6)) is close to zero. The next theorem says that the distribution of $\tilde X_{k,n,p}=\sum_{i=1}^{k}\tilde\zeta_i^p/\sum_{i=1}^{n}\tilde\zeta_i^p$ is close to the Beta $B(\frac{k}{p},\frac{n-k}{p})$-distribution w.r.t. the uniform distance. In the sequel c denotes absolute constants which may be different in different places, and c(...) denotes quantities depending only on the arguments in the paranthesis.

<u>THEOREM</u> 4.5. For any $0<p<\infty$ and $(\tilde\zeta_i)_{i\in\mathbb{N}}$ i.i.d. with $E\tilde\zeta_i^p=1$ and

$$\tilde m_\delta:=E\tilde\zeta_1^{(2+\delta)p}<\infty\quad(\delta>0)$$

we have

$$\Delta:=\sup_{k,n}\rho(X_{k,n,p},\tilde X_{k,n,p})\le c(\delta,\tilde m_\delta,p)\,\rho^{\frac{\delta}{3(2+\delta)}}. \tag{4.23}$$

<u>Sketch of the proof:</u> The proof is based on the relationship between metrics and the choice of the "natural" metric ζ_2 (see (4.2)).

Claim 1. *(The regularity of the uniform distance under convolutions):*

$$\rho(X_{k,n,p},\tilde{X}_{k,n,p}) \le \rho(\sum_{i=1}^{k}\zeta_i^p,\ \sum_{i=1}^{k}\zeta_i^p) + \rho(\sum_{i=k+1}^{n}\zeta_i^p,\ \sum_{i=k+1}^{n}\zeta_i^p)$$
$$\le n\,\rho(\zeta_1,\zeta_1).$$

Claim 2. *(Estimate from above for the "traditional" metric ρ by the "appropriate" metric ζ_2):* Let $n > p$, $E\zeta_1^p = E\tilde{\zeta}_1^p$, $\sigma_p^2 = \mathrm{Var}\,(\zeta_1^p) < \infty$. Then

$$\rho(\sum_{i=1}^{n}\zeta_i^p,\ \sum_{i=1}^{n}\tilde{\zeta}_i^p) \le 3\sigma_p^{2/3}(2\pi(1-\frac{p}{n}))^{-1/3}\zeta_2^{1/3}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_i,\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{Z}_i),$$

where $Z_i := \frac{\zeta_i^p - 1}{\sigma}$, $\tilde{Z}_i := \frac{\tilde{\zeta}_i^p - E\tilde{\zeta}_i^p}{\sigma}$.

Claim 3. *(The metric ζ_2 is "ideal" of order* 2 (see (4.10)):

$$\zeta_2(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_i,\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{Z}_i) \le \zeta_2(Z_1,\tilde{Z}_1).$$

Claim 4. *(Estimate from below of ρ by ζ_2).* If $m_\delta < \infty$, then

$$\zeta_2(Z_1,\tilde{Z}_1) \le c(\delta,\tilde{m}_\delta,p)\,\rho^{\frac{\delta}{2+\delta}}.$$

Combining all claims we get the required inequality. □

In the case $p = \infty$, we in fact change the summation scheme of i.i.d. r.v.'s (see Claim 3) with the maxima scheme of i.i.d. r.v.'s. In this case the "ideal" metric will be the weighted uniform distance

(4.24) $$\rho_\alpha(X,Y) = \sup_{x>0} x^\alpha |F_X(x) - F_Y(x)|,$$

[128], [129], [130], [78], [23]. Clearly, for any nonnegative independent r.v.'s $\{X_i\}$ and $\{Y_i\}$ and positive constants c_i, similarly to (4.10) we have

(4.25) $$\rho_\alpha(\bigvee_{i=1}^{n} c_iX_i,\ \bigvee_{i=1}^{n} c_iY_i) \le \sum_{i=1}^{n} c_i^\alpha \rho_\alpha(X_i,Y_i).$$

For the necessary inequalities between ρ and ρ_α we refer to [129], [23], [78].

One may want to obtain stronger results than in Theorem 4.3 by examining the deviation between $X_{k,n,p}$ and $\tilde{X}_{k,n,p}$ in terms of the total variation distance

(4.26) $$\sigma(X,Y) = \sigma(Pr_X,Pr_Y) = \sup_{A\in\mathfrak{B}(\mathbb{R})} |Pr(X\in A) - Pr(Y\in A)|.$$

THEOREM 4.6. For $0 < p < \infty$ the following holds:

$$\sigma(X_{n,k,p}, \tilde{X}_{n,k,p}) \le \sigma\Big(\sum_{i=1}^{k} \zeta_i^p, \sum_{i=1}^{k} \tilde{\zeta}_i^p\Big) + \sigma\Big(\sum_{i=k+1}^{n} \zeta_i^p, \sum_{i=k+1}^{n} \tilde{\zeta}_i^p\Big).$$

If $E\zeta_i^{pj} = E\tilde{\zeta}_i^{pj} = EN_i^j$, $i = 1,2$, $j = 1,2$, for some independent normal r.v.'s N_i and if for some $r > 2$ the pseudomoments

$$\upsilon_r = \upsilon_r(\zeta_i, N_i) = \int |x|^r |F_{\zeta_i^p}(x) - F_{N_i}(x)|\, dx \le a < \infty,$$

then

$$\sup_{k \le n} \sigma(X_{k,n,p}, \tilde{X}_{k,n,p}) \le A(a) n^{-1/2}. \tag{4.27}$$ □

The proof is similar to that of Theorem 4.4, but the "ideal metric is different", here we need the so-called *ideal smoothing metric of order r*

$$v_r(X,Y) = \sup_{h>0} h^r \sigma(X + hZ, Y + hZ), \tag{4.28}$$

where Z is $N(0,1)$-distributed and independent of X and Y ([92], [87], [97]). One can easily check that

$$v_r\Big(\sum_{i=1}^{n} c_i X_i, \sum_{i=1}^{n} c_i Y_i\Big) \le \sum_{i=1}^{n} |c_i|^r v_r(X_i, Y_i), \tag{4.29}$$

where $\{X_i\}$ and $\{Y_i\}$ are independent (cf. (4.10)).

Next we shall use these "characterization" results to study the stability in de Finetti's theorem. In the paper of Diaconis and Freedman [26] two "continuous" examples of de Finetti theorems are considered: Let ζ be chosen at random on the surface of the 2-sphere $O_{2,n} := \{x \in \mathbb{R}^n : \Sigma x_i^2 = n\}$ (resp. the simplex $S_{1,n} := \{x \in \mathbb{R}_+^n : \Sigma x_i = n\}$). Then $\zeta_1, \ldots, \zeta_k$ are, for k fixed, in the limit as $n \to \infty$ independent standard normals (resp. exponentials). Diaconis and Freedman obtained a right order bound on the variation distance between the law of $(\zeta_1, \ldots, \zeta_k)$ and the law of k independent standard normals (resp. exponentials). We will extend Diaconis' and Freedman's results considering ζ being chosen "at random" on the surface of the "p-sphere" $O_{p,n} = \{x \in \mathbb{R}^n : \sum_{i=1}^{n} |x_i|^p = n\}$ for $0 < p < \infty$ and $O_{\infty,n} = \{x \in \mathbb{R}^n, \bigvee_{i=1}^{n} |x_i| = n\}$. By this random choice on the surface we mean the measure (up to normalization) arising from a desintegration of the Lebesgue measure w.r.t. the surface $\{x \in \mathbb{R}^n ; \Sigma |x_i|^p = t\}$ parametrized by t. This uniform distribution is relevant for applications to physics in connection with the theorem of Liouville.

For $p=1,2,\infty$ this is identical with the geometric surface measure. Without loss of generality we consider the positive parts of $O_{p,n}$ namely $S_{p,n} = \{x \in O_{p,n};\ x_i \geq 0\}$. We start with the case $p=\infty$.

Let $\zeta_1,\dots,\zeta_n$ be i.i.d. uniformly $(0,1)$-distributed, then $(\zeta_1,\dots,\zeta_n)$ is conditionally given $\bigvee_{i=1}^{n} \zeta_i = s$ uniform on $S_{\infty,s,n} := \{x \in \mathbb{R}_+^n;\ \bigvee_{i=1}^{n} x_i = s\}$.

Let $P_\sigma^{n,\infty}$ for $\sigma > 0$ be the law of $(\sigma\zeta_1,\dots,\sigma\zeta_n)$ and let $Q_{n,s,k}^{(\infty)}$ be the law of $(\eta_1,\dots,\eta_k)$, where $\eta = (\eta_1,\dots,\eta_n)$ is uniform on $S_{\infty,s,n}$. In the next theorem we shall evaluate the deviation between $Q_{n,s,k}^{(\infty)}$ and $P_s^{k,\infty}$ in terms of the total variation distance $\sigma(Q_{n,s,k}^{(\infty)}, P_s^{k,\infty}) := \sup_{A \in \mathfrak{B}^k} |Q_{n,s,k}^{(\infty)}(A) - P_s^{k,\infty}(A)|$, $\mathfrak{B}^k$ being the Borel sets in $\mathbb{R}^k$.

<u>THEOREM</u> 4.7. $\sigma(Q_{nsk}^{(\infty)}, P_s^{k,\infty}) = \frac{k}{n}$. □

Let C_n be the class of distributions of $X = (X_1,\dots,X_n)$ on $\mathbb{R}_+^n$ which share **with the i.i.d. uniforms the property that given $M := \bigvee_{i=1}^{n} X_i = s$, the conditional** joint distribution of X is uniform on $S_{\infty,s,n}$.

Clearly $P_\sigma^{n,\infty} \in C_n$; so is $P_{\mu,n} = \int P_\sigma^{n,\infty}\mu(d\sigma)$ for any probability μ on $(0,\infty)$. As a consequence of Theorem 4.7 we get the finite form of the de Finetti-Theorem:

<u>COROLLARY</u> 4.1. If $P \in C_n$, then there is a μ such that for all $k < n$, $\sigma(P_k, P_{\mu k}) \leq \frac{k}{n}$, where P_k is the P-law of the first k-coordinates $(X_1,\dots,X_k)$.

In particular, one gets the infinite de Finetti-type characterization of scale mixtures of i.i.d. uniform variables (cf. Diaconis and Freedman [25], Example 2.5).

<u>COROLLARY</u> 4.2. Let P be a probability on $\mathbb{R}_+^\infty$ with P_n being the P-law of the first n coordinates. Then P is a uniform scale mixture of i.i.d. uniform variables, if and only if $P_n \in C_n$ for every n. □

Following the same idea we may consider the case $p \in (0,\infty)$. While for the case $p=\infty$ we use the stability results for maxima of i.i.d. r.v.'s in the case $p \in (0,\infty)$ we apply Theorem 4.3 and its stability versions, for details see [94].

THEOREM 4.8. For $0 < p < \infty$

(4.30) $$\sigma(Q_{nsk}^{(p)}, P^{k,p}_{(\frac{s}{n})^{1/p}}) \le \frac{1}{n-k-p}\,[1 + \frac{3}{2}k + (\frac{e^{1/12}}{12} + 1)p + (\frac{k}{2} + \frac{pe^{1/12}}{12}(1+\frac{k}{2}))\frac{1}{n} + \frac{pe^{1/12}}{24}\frac{1}{n^2}].$$ □

The order in (4.30) is essentially $\frac{k}{n}$. This result is asymptotically sharp as follows from

THEOREM 4.9. For $0 < p < \infty$, $\frac{k}{n} \to 0$ holds:

(4.31) $$\sup_{s>0} \sigma(Q_{nsk}^{(p)}, P^{k,p}_{(\frac{s}{n})^{1/p}}) = \frac{1}{4}E|1 - N_{0,1}^2|\frac{k}{n} + o(\frac{k}{n}),$$

where $N_{0,1}$ is a standard normal random variable.

EXAMPLE 4.4. *Approximation by compound Poisson distributions*
A famous problem of Kolmogorov on the uniform rate of approximation of sums of i.i.d. r.v.'s w.r.t. the class of infinitely divisible distributions was solved finally by Arak in 1981. While the optimal uniform approximation rate w.r.t. the "usual" accompanying laws is of order $n^{-1/3}$, the optimal order is $n^{-2/3}$. But the problem to construct the best approximations (in particular compound Poisson approximations) is still unsolved.

Let $X_1,\dots,X_n$ be independent, real values r.v.'s with df F_i, $1 \le i \le n$, of the form

(4.32) $$F_i = (1-p_i)E_o + p_i V_i, \quad 0 \le p_i \le 1,$$

where E_o is the one point df concentrated at zero and V_i is any df on $\mathbb{R}^1$. For the sum $S_n = \sum_{i=1}^n X_i$ the usual accompanying approximation by a compound Poisson distributed random variable is given by the parameters

(4.33) $$\mu = \sum_{i=1}^n p_i, \quad V = \sum_{i=1}^n \frac{p_i}{\mu} V_i$$

of the compound Poisson distributed random variable $\tilde{S}_n$. For this choice, which is very common in risk theory, one knows several bounds for the uniform and total variation metric, in particular:

(4.34) $$\sigma(S_n, \tilde{S}_n) \le \sum_{j=1}^n p_j^2,$$

(4.35) $$\rho(S_n, \tilde{S}_n) \le c \max_{1\le j\le n} p_j.$$

In the i.i.d. case $F_i = F$, $\forall i$

(4.36) $$\rho(S_n, \tilde{S}_n) \le c\, n^{-1/3}$$

(for relevant references cf. [99]).

Motivated by risk theory a natural metric for this approximation problem is in terms of the stop-loss distance of order s

(4.37) $$\begin{aligned} d_s(X,Y) &= \sup_t \left| \int_0^\infty \frac{(x-t)^s}{s!}\, d(F_X(x) - F_Y(x)) \right| \\ &= \sup_t \frac{1}{s!} \left| E(X-t)_+^s - E(Y-t)_+^s \right|, \; s \in \mathbb{N}. \end{aligned}$$

It turns out that d_s has good topological and metric properties (cf. [99]) allowing e.g. to derive a Berry-Esseen theorem for d_s in the case of i.i.d. r.v.'s. W.r.t. the problem of approximation by compound Poisson distributions a different choice of parameters than those in (4.33) leads to better and more stable approximtion results. Let C_i be r.v.'s with distributions V_i, $a_i = EC_i$, $b_i = EC_i^2$, define:

(4.38) $$\mu_i = \frac{p_i b_i}{b_i - p_i a_i^2}, \quad u_i = \frac{p_i}{\mu_i}$$

and consider the choice of parameters of the compound Poisson distribution

(4.39) $$\mu = \sum_{i=1}^n \mu_i, \quad V = \sum_{i=1}^n \frac{\mu_i}{\mu} F_{u_i C_i}.$$

Then we obtain that:

(4.40) $$ES_n = E\tilde{S}_n \quad \text{and} \quad \operatorname{Var} S_n = \operatorname{Var} \tilde{S}_n$$

in contrast to the choice in (4.34) where Var (S_n) is smaller than Var $(\tilde{S}_n)$. For the approximation which turns out to be very good in simulations, the following bounds have been established in Rachev and Rüschendorf [99].

THEOREM 4.10. For the choice of parameters as in (4.39) holds:

a) (4.41) $$d_1(S_n, \tilde{S}_n) \le \frac{4}{\sqrt{\pi}} \sqrt{\sum_{i=1}^n p_i b_i}\;;$$

b) (4.42) $$d_2(S_n, \tilde{S}_n) \le \sum_{i=1}^n p_i b_i.$$

If $C_i \ge 0$, then

c) (4.43) $$d_1(S_n, \tilde{S}_n) \le \sum_{i=1}^n p_i^2 \tau_i,$$

where $\tau_i = a_i + \Delta_i v_i + \max(\Delta_i a_i v_i, 2a_i v_i + 1 + v_i p_i)$

$$v_i = a_i^2 / b_i \le 1, \ \Delta_i := p_i v_i \le 1.$$

d) (4.44) $$d_2(S_n, \tilde{S}_n) \le \frac{1}{2} \sum_{i=1}^n p_i^2 \tau_i^*,$$

$$\tau_i^* = \frac{3}{2} b_i + 2a_i^2 + \frac{1}{2} \tilde{\Delta}_i a_i^2 + p_i \frac{a_i^4}{4b_i}, \ \tilde{\Delta}_i := p_i \frac{a_i^2}{b_i} \le 1,$$

(4.45) $$d_1(S_n, \tilde{S}_n) \le \frac{4}{\sqrt{\pi}} \sqrt{\sum_{i=1}^n p_i^2 \tau_i^*} .$$

In the proof of Theorem 4.10 we use some relations of the d_s metrics to other metrics and apply e.g. the following inequality:

(4.46) $$d_1(X,Y) \le \frac{4}{\sqrt{\pi}} \sqrt{d_2(X,Y)},$$

which is a consequence of a smoothing inequality for d_s. Note that for the usual choice of the parameters as in (4.33) one can not even assure the finiteness of $d_2(S_n, \tilde{S}_n)$. For the usual choice of the parameters a bound for $d_1(S_n, \tilde{S}_n)$ of a similar type as in (4.43) has been given before by Gerber [42].

EXAMPLE 4.5. *Approximation of dependent sequences of random variables.* One method of proving central limit theorems resp. invariance principles for dependent sequences is to approximate these sequences in a first step by a sequence with simpler structure as e.g. by an independent or a martingale sequence (cf. Berkes and Philipp [9], Eberlein [31], [32], the book of Philipp and Stout (cf. [108])). We shall consider two situations of this type.
In the first example we approximate a weakly dependent (φ-mixing) sequence by an independent sequence. The notion of φ-mixing dependence allows to give a natural formulation of the approximation in terms of the total variation distance (cf. (2.15)). If the weak dependence assumption is formulated in other terms (as e.g. the "very weak Bernoulli-condition"), then also different formulations of the approximation distance are more natural (cf. [108]). In the second example we consider the approximation by a martingale sequence. Since the martingale property is by definition in terms of conditional expectations it seems natural to use the minimizing property of expectations w.r.t. the L^2-distance and to formulate an approximation result in terms of the minimal L^2-distance. Some further general aspects

of the approximation by dependent sequences as e.g. Markov-sequences are discussed in [108]. A systematic investigation of the natural choice of metrics in these problems is still missing.

4.1. Weakly Dependent Sequences

Let $X = (X_k)$ be a sequence of random variables with values in universally measurable separable metric space (S_k, d_k) satisfying a φ-mixing condition

$$(4.47) \qquad |P(X_k \in A_k, (X_1,\dots,X_{k-1}) \in B_k) - P(X_k \in A_k) P((X_1,\dots,X_{k-1}) \in B_k)| \le \varphi_k P((X_1,\dots,X_{k-1}) \in B_k)$$

for $A_k \in \mathfrak{B}_k$, $B_k \in \mathfrak{B}_1 \otimes \dots \otimes \mathfrak{B}_k$, $n \in \mathbb{N}$. The following result improves upon a result in Berkes and Philipp [9].

THEOREM 4.11. [108]. There exist stochastic processes $Y = (Y_k)$, $Z = (Z_k)$ with

a) $X \overset{d}{\sim} Y$,

b) $\{Z_k\}$ independent, $Z_k \overset{d}{\sim} X_k$, $k \in \mathbb{N}$,

c) $P(Z_k \ne Y_k) \le \varphi_k$, $\forall k \in \mathbb{N}$.

4.2. Martingale Approximation

Let $X = (X_1,\dots,X_n)$ be any sequence of real r.v.'s in $L^2(P)$ and let $\mathfrak{A}_1 \subset \mathfrak{A}_2 \subset \dots \subset \mathfrak{A}_n \subset \mathfrak{A}$ be a sequence of σ-algebras such that X_k is $\mathfrak{A}_k$-measurable, $1 \le k \le n$. We consider the problem to find an optimal approximation of X by a martingale $Y = (Y_k, \mathfrak{A}_k)_{1 \le k \le n}$ w.r.t. the L^2-distance

$$(4.49) \qquad L^2(X,Y) = E\|X - Y\|^2 = \sum_{i=1}^{n} E(X_i - Y_i)^2 .$$

This problem was solved by Rüschendorf [108].

THEOREM 4.12. The optimal approximation of X by a martingale $(Y_k, \mathfrak{A}_k)$ is given by

$$(4.50) \qquad Y_1 = \frac{1}{n}\Big(X_1 + \sum_{\ell=2}^{n} E(X_\ell | \mathfrak{A}_1)\Big)$$

$$Y_k = \frac{1}{n-k+1}\Big[X_k + \sum_{\ell=k+1}^{n} \big(E(X_\ell | \mathfrak{A}_k) - E(X_\ell | \mathfrak{A}_{k-1})\big)\Big] + Y_{k-1}.$$

□

5. The Structure of "Ideal" Metrics

In the previous sections we considered several examples of "natural" ("ideal") metrics. Since almost any approximation problem has its own natural metric, the classification problem for ideal metrics is a difficult task. Zolotarev [122], [128] considered the "ideal" metrics for the basic schemes in probability theory, summation and maxima of i.i.d. r.v.'s. Ideal metrics for noncommutative operation between random elements (as random motions) are considered in [53], [97].

Further we present a simple classification of ideal metrics for summation and maxima of i.i.d. r.v.'s.

Let $(U, \| \|)$ be a separable Banach space with norm $\| \|$ and Borel σ-algebra $\mathfrak{B} = \mathfrak{B}(U)$ and let $\mathfrak{X} = \mathfrak{X}(U)$ be the set of all random variables on a non-atomic probability space $(\Omega, \mathfrak{A}, Pr)$ with values in U. Then the set $\mathfrak{P} = \mathfrak{P}(U)$ of all distributions $\{Pr_X;\ X \in \mathfrak{X}\}$ coincide with the set $M_1(U, \mathfrak{B})$ of all probability measures on $(U, \mathfrak{B})$. A function $\mu: \mathfrak{X} \times \mathfrak{X} \to [0,\infty]$ is called a *probability metric* (p metric) (cf. [122], p. 374, [91], [93]) if for $X,Y,Z \in \mathfrak{X}$

$$
(5.1)\qquad
\begin{aligned}
&1.\ Pr(X = Y) = 1 \to \mu(X,Y) = 0,\\
&2.\ \mu(X,Y) = \mu(Y,X),\\
&3.\ \mu(X,Z) \le \mu(X,Y) + \mu(Y,Z).
\end{aligned}
$$

μ is called a *simple metric* if $X_1 \overset{d}{=} X_2$, $Y_1 \overset{d}{=} Y_2$ implies $\mu(X_1,X_2) = \mu(Y_1,Y_2)$ and compound otherwise. A simple metric induces a (usual semi-) metric $\mu: \mathfrak{P}(U) \times \mathfrak{P}(U) \to [0,\infty]$ and vice versa.

Considering the rate of convergence problem for the CLT Zolotarev [121] introduced the notion of ideal metrics w.r.t. summation of i.i.d. r.v.'s. A metric μ is called a *compound* $(r,+)$*-idel metric* (ideal or order $r > 0$ w.r.t. summation), if and only if for all $X,Y,Z \in \mathfrak{X}$, $c \in \mathbb{R}^1$, the following holds:

$$
(5.2)\qquad
\begin{aligned}
&1.\ \mu(X+Z,\ Y+Z) \le \mu(X,Y),\\
&2.\ \mu(cX,cY) = |c|^r \mu(X,Y).
\end{aligned}
$$

μ is called a *simple* $(r,+)$*-ideal metric* if 1. is satisfied for any Z independent of X and Y.

A consequence of the (r,+)-ideality of μ is the estimate:

(5.3) $$\mu(\sum_{j=1}^{n} c_j X_j, \sum_{j=1}^{n} c_j Y_j) \le \sum_{j=1}^{n} |c_j|^r \mu(X_j, Y_j)$$

for any $c_j \in \mathbb{R}^1$ and any r.v.'s (X_j), (Y_j). If μ is a simple (r,+)-ideal metric, then $\{X_1,\dots,X_n\}$ as well as $\{Y_1,\dots,Y_n\}$ are supposed to be independent r.v.'s. In particular, if $X_1, X_2, \dots$ are i.i.d. r.v.'s and $Y_{(\alpha)}$ has a strictly symmetric stable distribution with parameter $\alpha \in (0,2]$ and μ is a simple (r,+)-ideal metric of order $r > \alpha$, then one gets from (5.3)

(5.4) $$\mu(n^{-1/\alpha} \sum_{i=1}^{n} X_i, Y_{(\alpha)}) \le n^{1-\frac{r}{\alpha}} \mu(X_1, Y_{(\alpha)}),$$

which gives a precise estimate in the CLT under the only assumption that $\mu(X_1, Y_{(\alpha)}) < \infty$. (Under additional assumptions one can improve the order of (5.4) using the minimal metric ℓ_p, see (2.12), [98]).

In several Banach spaces (e.g. in Banach function spaces) one has a natural maximum operation $x \vee y$. W.r.t. the operation $\vee$ one defines similarly the notion of *compound and of simple (r,v)-ideal* metrics assuming condition 2. in (5.2) only for positive c. Especially, if μ is a simple (r,v)-ideal metric on $\mathbb{R}^1$ and if $Z_{(\alpha)}$ is a α-max-stable r.v. on $\mathbb{R}^1$ (i.e. $F_{Z_{(\alpha)}}(x) = \exp\{-x^{-\alpha}\}$, $x \ge 0$), then

(5.5) $$\mu(n^{-1/\alpha} \bigvee_{i=1}^{n} X_i, Z_{(\alpha)}) \le n^{1-\frac{r}{\alpha}} \mu(X_1, Z_{(\alpha)})$$

for any i.i.d. r.v.'s X_i; for example $\mu = \ell_p$, $r = \min(1,p)$.

Further, we shall construct some ideal metrics for summation and for maxima and discuss the problem formulated by Zolotarev [127], p. 300 to construct metrics which are ideal w.r.t. both operatons simultaneously. As is immediately clear from (5.3) - (5.5) one gets as a consequence, rate of convergence results in the CLT for sums and maxima of i.i.d. r.v.'s. It will be interesting to compare the new results with clasicial results in terms of e.g. the Prohorov distance and also w.r.t. the assumptions in these theorems. We will point out the importance of the L_p-metrics (see (2.8)) in these kinds of problems and especially obtain an improvement of Zolotarev's classical estimate of the rate of convergence w.r.t. the Prohorov distance for $1 \le \alpha < 2$. We will show that Zolotarev's problem has essentially a negative answer. However, in the range $0 < \alpha < 2$, the minimal metrics $\ell_p = \hat{L}_p$ (in spite of being only ideal of order min (1,p)) behave like "doubly ideal" metrics of order $r = 1 + \alpha - \alpha/p \ge 1$ for $0 < \alpha \le p \le 2$. For the detailed proofs, see Rachev and Rüschendorf [98].

5.1. Ideal Metrics and Rate of Convergence for Summation

Recall the definition of L_p-metrics and their minimal metrics $\ell_p = \hat{L}_p$, see (2.18), (2.19). For $X, Y \in \mathfrak{X}(U) = \mathfrak{X}$

(5.6)
$$L_p(X,Y) = (E\|X - Y\|^p)^{\min(1,1/p)},\ 0 < p < \infty,$$
$$L_\infty(X,Y) = \text{ess sup } \|X - Y\|,$$

(5.7)
$$\ell_p(X,Y) = \inf\{L_p(\tilde{X},\tilde{Y});\ \tilde{X} \overset{d}{=} X,\ \tilde{Y} \overset{d}{=} Y\},\ 0 < p \le \infty.$$

Note that $\ell_p(X,Y) < \infty$ $(0 < p \le \infty)$ does not imply the finiteness of p-th moments of $\|X\|$ and $\|Y\|$; for example on $\mathbb{R}^1$ a sufficient condition for $\ell_p(X,Y) < \infty$ $(1 \le p < \infty)$ is $\varkappa_p(X,Y) := \int |x|^{p-1} |F_X(x) - F_Y(x)| dx < \infty$. The advantage of exploring the difference moment condition $\varkappa_p(X,Y) < \infty$ in Berry-Esseen type estimates was demonstrated by Hall [46]. Since L_p is a compound $(r,+)$-ideal metric with $r = \min(p,1)$, ℓ_p is a simple $(r,+)$-ideal metric [122], [91]. Therefore, from (5.4) one obtains for i.i.d. r.v.'s, $\{X_i\}$ and for $Y_{(\alpha)}$, the α-stable r.v., the estimate:

(5.8)
$$\ell_p\Big(n^{-1/\alpha} \sum_{i=1}^n X_i, Y_{(\alpha)}\Big) \le n^{1-\frac{r}{\alpha}} \ell_p(X_1, Y_{(\alpha)}),$$

which is useful only for $0 < \alpha < p \le 1$.

In the following remark we will discuss the results obtained by means of Zolotarev's ideal metric ζ_r. (Recall that ζ_2 for $U = \mathbb{R}$ was defined by (4.2).)

REMARK 5.1.

a) It is easy to see that there is *no* nontrivial compound $(r,+)$-ideal metric μ, when $r > 1$. Since the compound $(r,+)$ ideality would imply $\mu(X,Y) = \mu\left(\frac{X+\dots+X}{n}, \frac{Y+\dots+Y}{n}\right) \le n^{1-r}\mu(X,Y)$, $\forall n \in \mathbb{N}$, i.e. $\mu(X,Y) \in \{0,\infty\}$ for all $X, Y \in \mathfrak{X}(U)$, see the discussion in [54].

b) Zolotarev [121] - [127] found a simple $(r,+)$-ideal metric of any order $r > 1$, namely: If $r = m + \alpha$, $0 < \alpha \geq 1$, $m \in \mathbb{N}$, then:

$$(5.9) \qquad \zeta_r(X,Y) = \sup\{|E(f(X) - f(Y))|;\ |f^{(m)}(x) - f^{(m)}(y)| \leq |x - y|^{\alpha}\},$$

$f^{(m)}(x)$ denoting the Fréchet-derivative of order m, see (4.2) for $r = 2$, $U = \mathbb{R}$, $\|\ \| = |\ |$.

We next show that the minimal metric ℓ_p, in spite of being only a simple $(r_p,+)$-ideal metric, $r_p = \min(1,p)$, it acts as an ideal $(r,+)$ metric of order $r = 1 + \alpha - \alpha/p$ for $0 < \alpha \leq p \leq 2$. We formulate this result for Banach spaces U of type p (cf. [52], [120]).

THEOREM 5.1. If U is of type p, $1 \leq p \leq 2$ and $0 < \alpha < p \leq 2$, then for any i.i.d. r.v.'s $X_1,\dots,X_n \in \mathfrak{X}(U)$ with $E(X_i - Y_{(\alpha)}) = 0$ and for the strictly symmetric stable r.v. $Y_{(\alpha)}$ the following holds:

$$(5.10) \qquad \ell_p(n^{-1/\alpha} \sum_{i=1}^{n} X_i, Y_{(\alpha)}) \leq B_p^{1/p}\, n^{1/p-1/\alpha}\, \ell_p(X_1, Y_{(\alpha)}),$$

where $B_1 = 1$ and $B_p = 18\,p^{3/2}/(p-1)^{1/2}$ for $1 < p \leq 2$.

From Strassen-Dudley's representation of the Prohorov distance (see Section 2) one obtains the relation $\pi^{p+1} \leq (\ell_p)^p$, $\pi := \pi_1$. The last inequality implies the following corollary.

COROLLARY 5.1. Under the assumptions of Theorem 5.1, for $1 \leq p \leq 2$, $0 < \alpha < p \leq 2$ we have

$$(5.11) \qquad \pi(n^{-1/\alpha} \sum_{i=1}^{n} X_i, Y_{(\alpha)} \leq B_p^{\frac{1}{p+1}}\, n^{\frac{1}{p+1}(1-\frac{p}{\alpha})}\, \ell_p(X_1, Y_{(\alpha)})^{\frac{p}{p+1}}.$$

Theorem 5.1 and Corollary 5.1 are another example of the two stage approach to an approximation problem described in the previous section. We first solve the approximation problem in terms of the "natural" metric ℓ_p (see (5.10)) and then if we want to see the result in terms of the "classic" metric π we construct inequalities. Similarly one can use ζ_r instead of ℓ_p and then compare ζ_r with π, [121], [8], but the results cited above improve those in [121] and [8], see also [98].

Open Problem 5.1. Find "ideal metrics" for rate of cnvergence of "stochastic convolutions", see [10], [43], [63], [64], [116], [117], [79], [62],

[118]. Clearly, such ideal metrics will supply the right order estimates in the CLT for "stochastic" convolutions.

5.2. Ideal Metrics and Rate of Convergence for Maxima

For the maxima of r.v.'s several simple (r,v)-ideal metrics are known for any $r > 0$, implying by (5.5) the rate of convergence of order $1 - r/\alpha$ (cf. [129], [78], [23]). In the following example we construct for any $r > 0$ a compound (r,v)-ideal metric. This shows an essential difference between summation and maxima of r.v.'s.

EXAMPLE 5.1. *(A compound (r,v)-ideal metric)*
For $U = \mathbb{R}^1$ and any $0 < p \le \infty$ define for $X, Y \in \mathfrak{X}(\mathbb{R}^1)$

$$\Delta_{r,p}(X,Y) = \left(\int_{-\infty}^{\infty} \varphi^p_{X,Y}(x)\,|x|^{rp-1}dx\right)^q \tag{5.12}$$

and

$$\Delta_{r,\infty}(X,Y) = \sup_{x\in\mathbb{R}^1} |x|^r \varphi_{X,Y}(x),$$

where $q = \min(1, \frac{1}{p})$ and $\varphi_{X,Y}(x) = \Pr(X \le x < Y) + \Pr(Y \le x < X)$. One can easily check that $\Delta_{r,p}$ is a compound $(r \min(p,1),v)$-ideal metric for $0 < p \le \infty$, $0 < r < \infty$. The corresponding minimal metric $\hat{\Delta}_{r,p}$ (see (2.13)) is a simple $(r \min(p,1),v)$-ideal metric which one can check by the representation

$$\hat{\Delta}_{r,p}(X,Y) = \left(\int_{-\infty}^{\infty} |x|^{rp-1}|F_X(x) - F_Y(x)|^p dx\right)^q,\ 0 < p < \infty, \tag{5.13}$$

$$\hat{\Delta}_{r,\infty}(X,Y) = \sup_{x\in\mathbb{R}} |x|^r |F_X(x) - F_Y(x)|$$

(cf. [83], [86]).

From (5.5) one obtains for a simple (r,v)-ideal metric μ that $\mu(X_1, Z_{(\alpha)}) < \infty$ implies $\mu(n^{-1/\alpha} \bigvee_{i=1}^{n} X_i, Z_{(\alpha)}) \le n^{1-r/\alpha} \mu(X_1, Z_{(\alpha)})$.

For $\mu = \hat{\Delta}_{r,\infty}$ it was shown by Omey and Rachev [78] that also the converse relation is correct, i.e. the rate in (5.5) is of right order.

We next want to investigate the properties of the L_p-metrics (cf. 5.6) w.r.t. maxima. We consider for $0 < \lambda \le \infty$ the Banach space $U = \Lambda_{\lambda,\mu} = \{x: (E,\mathfrak{E}) \times (\Omega,A) \to (\mathbb{R}^1,\mathfrak{B}^1);\ \|x\|_{\lambda,\mu} \le \infty\}$ where $(E,\mathfrak{E},\mu)$ is a measurable space and

$$\|x\|_{\lambda,\mu} = \begin{cases} (\int |x(t)|^{\lambda^*} d\mu(t))^{1/\lambda} & \text{for } 0 < \lambda < \infty \\ \text{ess sup } |x(t)| & \text{for } \lambda = \infty \end{cases}$$

with $\lambda^* = \max(\lambda,1)$ and define for $x,y \in U$, $x \vee y$ to be the pointwise maximum, $x \vee y(t) = x(t) \vee y(t)$, $t \in E$. For related limit results for α-max stable processes see [23]. In the case $\lambda = \infty$, $\lambda_{\infty,\mu}$ is not separable but since the ess sup $|X(t)|$ is measurable this does not cause difficulties.

The estimate (5.8) is interesting for $p \le 1$ only; for $1 < p \le \lambda < \infty$ one can improve it as follows:

THEOREM 5.2. [98] Let $1 \le p \le \lambda < \infty$, then for $X_1,\dots,X_n \in \mathfrak{X}(\Lambda_{\lambda,\mu})$ i.i.d. holds

$$\ell_p(n^{-1/\alpha} \bigvee_{i=1}^{n} X_i, Z_{(\alpha)}) \le n^{1/p-1/\alpha} \ell_p(X_1, Z_{(\alpha)}). \tag{5.14}$$

Comparing (5.14) with (5.8) we see that actually ℓ_p "acts" in this important case as a simple $(\alpha + 1 - \alpha/p, \vee)$-ideal metric.

Open Problem 5.2. Smith [113], Cohen [15], Resnick [102], and Balkema and de Haan [2] consider the univariate case $(X,X_1,X_2,\dots \in \mathfrak{X}(\mathbb{R}))$ of general normalized maxima

$$\rho(a_n \bigvee_{i=1}^{n} X_i - b_n, Y) \le c(X_1,Y)\Phi_{X_1}(n), \quad n = 1,2,\dots, \Phi_{X_1}(n) \to 0.$$

In order to extend these type results to the multivariate case $(X,X_1,X_2,\dots \in X(\Lambda_{\lambda,\mu}))$ using the theory of probability metrics, one needs to generalize (5.5) by determining metrics μ_Φ and $\overline{\mu}_\Phi$ in $X(\Lambda_{\lambda,\mu})$ such that for any $X_1,X_2,Y \in \mathfrak{X}(\Lambda_{\lambda,\mu})$ and $c > 0$

$$\mu_\Phi(c(X_1 \vee Y), c(X_2 \vee Y)) \le \Phi(c)\overline{\mu}_\Phi(X_1,X_2),$$

where $\Phi: [0,\infty) \to [0,\infty)$ is a suitably chosen strictly increasing continuous function, $\Phi(0) = 0$.

5.3. Doubly Ideal Metrics

We now investigate the question of the existence and construction of doubly ideal metrics posed by Zolotarev [127]. As we have shown in Section 5.1 and 5.2, ℓ_p are ideal metrics of order $\min(1,p) \le 1$ for both operations simultaneously. Let U be a Banach space with maximum operation $\vee$.

(Doubly ideal metrics). A probability metric μ on $X(U)$ is called

a) (r,I)-ideal, if μ is compound (r,+)-ideal and compound (r,v)-ideal;

b) (r,II)-ideal, if μ is compound (r,v)-ideal and simple (r,+)-ideal;

c) (r,III)-ideal, if μ is simple (r,v)-ideal and simple (r,+)-ideal.

L_p, $0 \le p \le \infty$, is an example of a (min (1,p),I)-ideal metric. we have seen in Section 5.1 that there does not exist a (r,I)-ideal metric for $r > 1$. ℓ_p is a (r,III)-ideal metric of order $r = \min(1,p)$. We now show that Zolotarev's question on the existence of a (r,II) or a (r,III)-ideal metric has essentially a negative answer.

<u>THEOREM</u> 5.3. [98] Let $r > 1$ and let μ be a (r,III)-ideal metric in $\mathfrak{P}(\mathbb{R}^1)$ and assume that μ satisfies the following regularity conditions:

C 1. If X_n (resp. Y_n) converges weakly to a constant a (resp. b), then

$$\overline{\lim_{n\to\infty}}\ \mu(X_n, Y_n) \ge \mu(a,b);$$

C 2. $\mu(a,b) = 0 \Leftrightarrow a = b$.

Then for any integrable $X, Y \in \mathfrak{X}(\mathbb{R})$ holds: $\mu(X,Y) \in \{0,\infty\}$. □

<u>REMARK</u> 5.2. Condition C 1 seems to be quite natural. Let e.g. F be a class of non-negative lower semicontinuous functions on $\mathbb{R}^2$ and $\varphi: [0,\infty) \to [0,\infty)$ be continuous, nondecreasing. Define the minimal metric

$$\mu(X,Y) = \inf\{\varphi(\sup_{f\in F} Ef(\tilde{X},\tilde{Y}));\ \tilde{X} \overset{d}{=} X,\ \tilde{Y} \overset{d}{=} Y\}.$$

Then μ is lower semicontinuous on $\mathfrak{X}(\mathbb{R}^1) \times \mathfrak{X}(\mathbb{R}^1)$, i.e. $(X_n, Y_n) \xrightarrow{w} (X,Y))$, implies $\underline{\lim}\ \mu(X_n, Y_n) \ge \mu(X,Y)$; so especially C 1 is fulfilled.

Nevertheless we shall show next that for $0 < \alpha \le 2$ the metrics ℓ_p for $1 < p \le 2$ "act" as (r,II)-ideal metrics in the rate of convergence problem for

$$\ell_p(Z_n, Z_n^*) \to 0,$$

where

$$Z_n = n^{-1/\alpha} \bigvee_{k=1}^{n} S_k,\ Z_n^* = n^{-1/\alpha} \bigvee_{k=1}^{n} S_k^*, \text{ and } S_k = \sum_{i=1}^{n} X_i,\ S_k^* = \sum_{i=1}^{n} X_i^*$$

are sums of i.i.d. r.v.'s. The order of ideality is $r = 2\alpha + 1 - \alpha/p > 2\alpha$ and, therefore, we obtain a rate of convergence $n^{2-r/\alpha}$.

We consider the case that $\{X_i\}$, $\{X_i^*\}$ are i.i.d. r.v.'s in $(U, \| \|) = (\ell_p, \| \|_p)$, where for $x = \{x^{(j)}\} \in \ell_p$, $\|x\|_p = (\sum_{j=1}^{n} |x^{(j)}|^p)^{1/p}$. For $x, y \in \ell_p$ we define $x \vee y = x^{(j)} \vee y^{(j)})$.

<u>THEOREM</u> 5.4. [98] Let $0 \le \alpha < p \le 2$, $1 \le p \le 2$ and $E(X_1 - X_1^*) = 0$, then under the conditions formulated above holds

$$\ell_p(Z_n, Z_n^*) \le (\frac{p}{p-1})^{1/p} B_p^{1/p} n^{\frac{1}{p}-\frac{1}{\alpha}} \ell_p(X_1, X_1^*).$$

In particular for the Prohorov metric π we have

$$\pi(Z_n, Z_n^*) \le (\frac{p}{p-1})^{\frac{1}{p+1}} B_p^{\frac{1}{p+1}} n^{\frac{1}{p+1}(1-\frac{p}{\alpha})} \ell_p^{\frac{p}{p+1}} (X_1, X_1^*). \qquad \square$$

Theorem 5.4 has some interesting consequences for the approximation of queuing systems (cf. [98]).
As we have seen there is no selection rule that choses the "natural" ("ideal", "appropriate") metric. The only way to find out what will be the "natural" metric for the given approximation problem is to know more about the properties of metrics and here the following basic research directions arise:

I. Description of the basic structure of metrics (semimetrics, uniformities) in the space of probability measures and random variables.

II. Analysis of the topologies in the space of probability measures, generated by different types of p-metrics. This analysis can be carried out with the help of compactness criteria for different metrics.

III. Characterization of p-metrics which are "natural" ("suitable", "ideal") w.r.t. the given approximation problem.

IV. Investigation of the main relationships between different types of metrics (semimetrics, uniformities, topologies) in the space of probability measrues and random variables.

V. Solution of the Kantorovich- resp. Fréchet-type optimization problems arising in the construction of minimal metrics or in transportation problems.

Acknowledgement. This research was supported by grants of the Deutsche Forschungsgemeinschaft (DFG) which is gratefully acknowledged. These grants made possible the cooperation of both authors. A main part of the work described in this paper was done during two visits of the authors at the "Institut für Mathematische Statistik", Münster, in 1988, resp. the "Department of Statistics and Applied Probability", Santa Barbara, in 1990, supported by the DFG.

REFERENCES

[1] Anderson, T. W.: An Introduction to Multivariate Statistical Analysis. Wiley & Sons, New York, 1984

[2] Balkema, A. A., and de Haan, L.: A convergence rate in extreme-value theory. Preprint, 1987

[3] Barlow, R. E., and Proschan, F.: Statistical of Reliability and Life Testing: Probability Models. Holt, Rinehart, and Winston, New York, 1975

[4] Barnes, R., and Hoffmann, J.: On transportation problems with upper bounds on leading rectangles. SIAM J. Algebraic Discrete Methods 6 (1985), 487 - 496

[5] Baxter, L., and Rachev, S. T.: A note on the stability of the estimation of the exponential distribution. Statist. & Prob. Letters 10 (1990), 37 - 41

[6] Baxter, L., and Rachev, S. T.: The stability of a characterization of the bivariate Marshall-Olkin distribution, 1990. To appear in: J. Math. Anal. and Appl.

[7] Bazaraa, M. S., and Jarvis, J. J.: Linear Programming and Network Flows. John Wiley & Sons, New York, 1977

[8] Bentkus, V. Yu., and Rachkauskas, A.: Estimates of the distance between sums of independent random elements in Banach spaces. Theor. Prob. Appl. 29 (1985), 50 - 65

[9] Berkes, I., and Philipp, W.: Approximation theorems for independent and weakly dependent random vectors. Ann. Prob. 7 (1979), 29 - 54

[10] Bochner, S.: Positive zonal functions on spheres. Proc. Nat. Acad. Sci. USA 40 (1954), 1141 - 1147

[11] Brown, M.: Approximating IMRL distributions by exponential distributions, with applications to first passage times. Ann. Prob. 11 (1983), 419 - 427

[12] Brown, M.: Exponential approximations for two classes of aging distributions. Ann. Prob. 12 (1984), 869 - 875

[13] Cambanis, S., Simons, G., and Stout, W.: Inequalities for Ek(X,Y) when the marginals are fixed. Z. Wahrsch. Verw. Geb. 36 (1976), 285 - 294

[14] Cambanis, S., and Simons, G.: Probability and expectation inequalities. Z. Wahrsch. Verw. Geb. 59 (1982), 1 - 25

[15] Cohen, J. P.: Convergence rate for the ultimate and penultimate approximation in extreme-value theory. Adv. Appl. Prob. 14 (1982), 833 - 854

[16] Cramer, H.: Mathematical Methods of Statistics. Princeton Univ. Press, New York, 1946

[17] D'Aristotile, A., Diaconis, P., and Freedman, D.: On merging of probabilities. Technical Report No. 301, Aug. 1988, Dept. of Statistics, Stanford Univ., California

[18] Daley, D. J.: Tight bounds on the exponential approximation of some aging distributions. Ann. Prob. 16 (1988), 414 - 423

[19] Dall'Aglio, G.: Sugli estremi dei momenti delle funzioni di ripartizione dopia. Ann. Scuola Normale Superiore Di Pisa, ser. C 1. Sci. 3.1 (1956), 33 - 74

[20] Dall'Aglio, G.: Les fonctions extremes de la classe de Fréchet a 3 dimensions. Publ. Inst. Stat. Univ. Parix IX (1960), 175 - 188

[21] Dall'Aglio, G.: Fréchet classes and compatibility of distribution functions. Symp. Math. 9 (1972), 131 - 150

[22] De Acosta, A.: Invariance principles in probability for triangle arrays of B-valued random vectors and some applications. Ann. Prob. 10 (1982), 346 - 373

[23] de Haan, L., and Rachev, S. T.: Estimates of the rate of convergence for max-stable processes. Ann. Prob. 17 (1989), 651 - 677

[24] Diaconis, P., and Freedman, D.: Partial exchangeability and sufficiency. In: Sufficiency and Statistics, Applications and New Directions, eds.: J. K. Ghosh and J. Roy., 1984, 205 - 236

[25] Diaconis, P., and Freedman, D.: A note on weak star uniformities. Technical report No. 39, Dept. of Statistics, Univ. of California, Berkeley, 1984

[26] Diaconis, P., and Freedman, D.: A dozen de Finetti-style results in search of a theory. Ann. Inst. Henri Poincare 23 (1987), 397 - 423

[27] Dimitrov, B., Khalil, Z., and Rachev, S. T.: A probabilistic approach to optimal quality usage. To appear in the special issue of "Computers and Mathematics with Applications" on "System-theoretic Methods in Economic Modelling", 1989

[28] Dudley, R. M.: Distances of probability measures and random variables. Ann. Math. Statist. 39 (1968), 1563 - 1572

[29] Dudley, R. M.: Probabilities and Metrices: Convergence of Laws on Metric Spaces, With a View of Statistical Testing. Aarhus Univ. Mat. Inst. Lecture Notes Series No. 45, Aarhus, 1976

[30] Dudley, R. M.: Real Analysis and Probability. Wadsworth and Brooks/ Cole, Math. Series, Pacific Grove, California, 1989

[31] Eberlein, E.: Strong approximation of very weak Bernoulli processes. Z. Wahrsch. Verw. Geb. 50 (1979), 119 - 133

[32] Eberlein, E.: Strong approximation of continuous time stochastic processes. J. Mult. Anal. 31 (1989), 220 - 235

[33] Fan, Ky: Entfernung zweier zufälliger Größen und die Konvergenz nach Wahrscheinlichkeit. Math. Z. 49 (1944), 681 - 683

[34] Fortét, R., and Mourier, E.: Convergence de la réparition empirique vers la reparition theoretique. Ann. Sci. Ecole Norm. Sup. 70 (1953), 267 - 285

[35] Fréchet, M.: Generalites sur les probabilites variable aleatoires. Gauthier-Villars, Paris, 1937

[36] Fréchet, M.: Sur des tableaux de correlation dont les marges sont donnees. Ann. Univ. Lyon 14 (1951), 57 - 77

[37] Fréchet, M.: Les tableaux de corrélation dont les marges et des bornes sont donnees. Ann. Univ. Lyon Sec. A 20 (1957), 13 - 31

[38] Fréchet, M.: Sur la distance de deux lois de probabilité. Publ. Inst. Statist. Univ. Paris 6 (1957), 185 - 198

[39] Gaffke, N., and Rüschendorf, L.: On a class of extremal problems in statistics. Math. Operat. Stat., Ser. Optimization 12 (1981), 123 - 135

[40] Gangolli, R.: Isotropic infinitely divisible measures on symmetric spaces. Acta Math. 111 (1964), 213 - 246

[41] Gelbrich, M.: On a formula for the L^2 Wasserstein metric between measures on Euclidean and Hilbert spaces. Preprint, 1988. 179, Sektion Mathematik der Humboldt-Universität zu Berlin

[42] Gerber, H.: Error bounds for the compound Poisson approximation. Insurance: Mathematics and Economics 3 (1984), 191 - 194

[43] Getoor, K. R.: Infinitely divisible probabilities on the hyperbolic plane. Pacific J. Math. 11 (1961), 81 - 115

[44] Gini, C.: Di una misura della dissomiglianza tra due gruppi di qualita e degli sue applicazioni allo studio delle relazioni statistiche. Atti de Reale Instituto Veneto di Sci, Lettera et Arti, LXXIV (1914), 185 - 213

[45] Gini, C.: La dissomiglianza. Metron. XXIV (1965), 85 - 215

[46] Givens, C. R., and Shortt, R. M.: A class of Wasserstein metrics for probability distributions. Michigan Math. J. 31 (1984), 231 - 240

[47] Grigorevski, N. N., and Shiganov, I. S.: On certain modifications of Dudley's metric (in Russian). Zapiski Nauchnyh Seminarov LOMI 41 (1976), 17 - 24

[48] Hall, P.: Two-sided bounds on a rate of convergence to a stable law. Z. Wahrsch. Verw. Geb. 57 (1981), 349 - 364

[49] Hennequin, P. L., and Tortrat, A.: Theorie des Probabilities et Quelques Applications. Masson, Paris, 1965

[50] Hoeffding, W.: Maßstabinvariante Korrelationstheorie. Sem. Math. Inst. Univ. Berlin 5 (1940), 181 - 233

[51] Hoffmann, A. J.: On simple linear programming problems, convexity. Proceedings of Symposia in Pure Mathematics 7, American Mathematical Society, Providence, RI, 1961, 317 - 327

[52] Hoffmann-Jorgensen, J.: Probability in Banach Spaces. In: LNM 598, pp. 1 - 187, Springer-Verlag, 1977

[53] Hohlov, Yu. S.: The estimation of the rate of convergence in the integral limit theorem in the Euclidean motion group. LNM 1233, pp. 1 - 10, Springer-Verlag, 1986

[54] Ignatov, Z. and Rachev, S. T.: Minimality of ideal probabilistic metrics. In: Stability Problems for Stochastic Models. Proceedings, Moscow, VNIISI, pp. 35 - 48 (in Russian), 1983; English transl., J. Soviet Math. 32 (1986), 595 - 608

[55] Kalashnikov, V. V., and Rachev, S. T.: Mathematical Methods for Construction of Queueing Moels. Nauka, Moscow (in Russian), 1988. English translation, Wadsworth & Brooks/Cole, California, 1990

[56] Kantorovich, L. V., and Rubinstein, G. Sh.: On the space of completely additive functions. Vestnik LGU, Ser. Mat., Mekh. i Astron. 7 (1958), 52 - 59 (in Russian)

[57] Kantorovich, L. V., and Akilov, G. P.: Functional Analysis. Nauka, Moscow (in Russian), 1984

[58] Karatzas, I.: Gittins indices in the dynamic allocation problem for diffusion processes. Ann. Prob. 12 (1984), 173 - 192

[59] Kellerer, H. G.: Duality theorems and probability metrics. In: Proc. 7th Brasov Conf. 1982, Bucuresti, 1984, 211 - 220

[60] Kellerer, H. G.: Duality theorems for marginal problems. Z. Wahrsch. Verw. Geb. 67 (1984), 399 - 432

[61] Kemperman, J. H. B.: On the role of duality in the theory of moments. In: Proc. Semi-Infinite Programming and Applicatios 1981, Lecture Notes in Economics and Math. Systems 215, pp. 63 - 92. Springer-Verlag, 1983

[62] Kendall, D. G.: Delphic semigroups, infinitely divisible regenerative phenomena, and arithmetic of p-functions. Z. Wahrsch. Verw. Geb. 9 (1968), 163 - 195

[63] Kennedy, M.: A stochastic process associated with the ultraspherical polynomials. Proc. Royal Irish. Acad., Sect. A, 61 (1961), 89 - 100

[64] Kingman, J. F.: Random walks with spherical symmetry. Acta Math. 109 (1963), 11 - 53

[65] Klefsjo, B.: The HNBUE and HNWUE classes of life distributions. Naval Research Logistics Quarterly 29 (1982), 331 - 344

[66] Knott, M., Smith, C. S.: On the optimal mapping of distributions. J. Optim. Th. Appl. 43 (1984), 39 - 49

[67] Kolmogorov, A. N.: Some recent work in the field of limit theorems in probability theory. Moscow Vestnik 10 (1953), 29 - 38

[68] Levin, V. L., and Rachev, S. T.: New dualty theorems for marginal problems with some applications in stochastics. In: Proc. Stability Problems for Stochastic Models, LNM 1412, 1990, 137 - 170

[69] Lévy, P.: Calcul des Probabilities. Gauthier-Villars, Paris, 1925

[70] Lévy, P.: Distance de deux variable aleatoires et distance de deux lois de probabilite. In: Generalites sur les probabilities variables aleatoires, by M. Frechet. Gauthier-Villars, Paris, 1937, 286 - 301

[71] Lukacs, E.: Stochastic Convergence. D. C. Heath and Co., Lexington, Mass., 1968

[72] Maejima, M., and Rachev, S. T.: An ideal metric and the rate of convergence to a self-similar process. Ann. Prob. 15 (1987), 708 - 727

[73] Marshall, A. W., and Olkin, I.: A multivariate exponential distribution. J. Amer. Stat. Assoc. 62 (1967), 30 - 44

[74] Monge, A.: Memoire sur la theorie des deblais et des remblais. Histoire de l'Academie des sciences de Paris, avec les Memoires de mathematique et de physique pour la meme annee, 1781, pp. 666 - 704, avec 2 pl.

[75] Neveu, J., and Dudley, R. M.: On Kantorovich-Rubinstein theorems. Preprint, 1980

[76] Obretenov, A., and Rachev, S. T.: Stability of some characterization properties of the exponential distribution. In: Stability Problems for Stochastic Models, Proceedings, Moscow, VNIISI, 1983, pp. 79 - 87 (in Russian); English transl. in J. Soviet Math. 32 (1986), 643 - 651

[77] Olkin, I., and Pukelsheim, F.: The distance between two random vectors with given dispersion matrices. Lin. Algebra Appl. 48 (1982), 257 - 263

[78] Omey, E., and Rachev, S. T.: On the rate of convergence in extreme value theory. Theorja Verojatnostei i Primenenia XXXII (1988), 601 - 607

[79] Partasarathy, K. R., Ranga, R. R., and Varadhan, S. R. S.: Probability distributions on locally compact abelian groups. Illinois J. Math. 7 (1963), 337 - 369

[80] Philipp, W., and Stout, W.: Almost sure invariance principles for sums of weakly dependent random variables. Am. Math. Soc. Memoir 161 (19)

[81] Prohorov, Yu. V.: Convergence of random processes and limit theorems in probability theory. Theory Prob. Appl. 1 (1956), 157 - 214

[82] Rachev, S. T.: Hausdorff metric structure of the space of probability measures. Zap. Nauchn. Sem. Leningrad. Otdel Mat. Inst. Steklov. (LOMI) 87 (1979), 87 - 103 (in Russian); English transl. J. Soviet Math. 17 (1981), 2218 - 2232

[83] Rachev, S. T.: On minimal metrics in a space of real-valued random variables. Soviet Math. Dokl. 23 (1981), 425 - 428

[84] Rachev, S. T.: Minimal metrics in the random variables space. In: Probability and Statistical Inference, Proceedings of the 2nd Pannonian Symp. Ed. by Grossman, M., et al., pp. 318 - 327. D. Reidel, Dodrecht, 1982

[85] Rachev, S. T.: Minimal metrics in the random variables space. Publ. Inst. Statist. Univ. Paris XXVII (1982), 27 - 47

[86] Rachev, S. T.: Minimal metrics in the real valued random variable space, LNM, pp. 172 - 180. Springer-Verlag, 1983

[87] Rachev, S. T., and Yukich, J.: Rate for CLT via new ideal metrics. Ann. Prob. 17 (1989), 775 - 788

[88] Rachev, S. T.: On a problem of Dudley. Soviet Math. Dokl. 29 (1984); 162 - 164

[89] Rachev, S. T.: On a class of minimal functionals in a space of probability measures. Theory Prob. Appl. 29 (1984), 41 - 49

[90] Rachev, S. T.: The Monge-Kantorovich mass transfer problem and its stochastic applications. Theroy Prob. Appl. 29 (1984), 647 - 676

[91] Rachev, S. T.: Extreme functionals in the space of probability measures. In: Proc. Stability Problems for Stochastic Models 1984, LNM 1155, 320 - 348. Springer-Verlag, 1985

[92] Rachev, S. T., and Ignatov, Z.: Ideal quadratic metrics. In: Stability Problems for Stochastic Models, Proceedings, Moscow, VNIISI, 119 - 128 (in Russian); English transl. J. Soviet Math. 35 (1986), 2396 - 23

[93] Rachev, S. T., and Shortt, R. M.: Classification problem for probability metrics. Proceedings of Conference in honor of Dorothy Maharam Stone, Univ. of Rochester, 1987. Contemporary Mathematics "Measure and Measurable Dynamics", AMS 1989, 221 - 262

[94] Rachev, S. T., and Rüschendorf, L.: Approximate independence of distributions on spheres and their stability properties. Preprint, 1988

[95] Rachev, S. T., and Taksar, M. I.: Kantorovich functionals in the space of probability measures and their applications in probability theory. Preprint, 1989

[96] Rachev, S. T., and Shortt, R. M.: Duality theorems for Kantorovich-Rubinstein and Wasserstein functions. Dissertationes Mathematicae 299 (1990)

[97] Rachev, S. T., and Yukich, J. E.: Smoothing metrics for measures on groups with applications to random motions. Preprint, 1989

[98] Rachev, S. T., and Rüschendorf, L.: Rate of convergence for sums and maxima and doubly ideal metrics. Preprint, 1989

[99] Rachev, S. T., and Rüschendorf, L.: Approximation of sums by compound Poisson distributions with respect to stop-loss distances. To apper in: J. Appl. Prob., 1990

[100] Rachev, S. T., Rüschendorf, L., and Schief, A.: On the construction of almost surely convergent random variables. Preprint series "Angewandte Mathematik und Informatik" 10/1988

[101] Rachev, S. T., Rüschendorf, L., and Schief, A.: Uniformities for convergence in distribution and stochastic convergence. Preprint, 1990

[102] Resnick, S.: Uniform rates of convergence to extreme-value distributions. In: Probability in Statistics: Essays in Honor of Franklin A. Graybill (J. Srivastava, ed.), North-Holland, Amsterdam, 1987

[103] Rockafellar, R. T.: Convex Analysis. Princeton Univ. Press, 1970

[104] Rolski, T.: Mean residual life. Z. Wahrsch. Verw. Geb. 33 (1975), 714 - 718

[105] Rüschendorf, L.: Inequalities for the expectation of Δ-monotone functions. Z. Wahrsch. Verw. Geb. 54 (1980), 341 - 349

[106] Rüschendorf, L.: Sharpness of Fréchet-bounds. Z. Wahrsch. Verw. Geb. 57 (1981), 293 - 302

[107] Rüschendorf, L.: Solution of a statistical optimization problem by rearrangement methods. Metrika 30 (1983), 55 - 62

[108] Rüschendorf, L.: The Wasserstein distance and approximation theorems. Z. Wahrsch. Verw. Geb. 70 (1985), 117 - 129

[109] Rüschendorf, L., and Rachev, S. T.: A characterization of random variables with minimum L^2-distance. J. Mult. Anal. 32 (1990), 48 - 54

[110] Rüschendorf, L.: Bounds for distributions with multivariate marginals. Preprint, 1989

[111] Rüschendorf, L.: Fréchet-bounds and their applications. Preprint, 1990

[112] Senatov, V. V.: Some lower estimates for the rate of convergence in the central limit theorem in Hilbert space. Sov. Math. Dokl. 23 (1981), 188 - 192

[113] Smith, R. L.: Uniform rates of convergence in extreme-value theory. Adv. Appl. Prob 14 (1982), 600 - 622

[114] Strassen, V.: The existence of probability measures with given marginals. Ann. Math. Statist. 36 (1965), 423 - 439

[115] Tchen, A. H.: Inequalities for distributions with given marginals. Ann. Prob. (1980), 814 - 827

[116] Urbanik, K.: Generalized convolutions. Studia Math. 23 (1964), 217 - 245

[117] Urbanik, K.: Remarks on B-stable probability distributions. Bul. Asad. Pol. Sci. Ser. Sci. Math. Astr. et Phys. 24 (1976), 783 - 787

[118] Volkovich, B. E.: Multivariate b-stable distributions and realization of generalized convolutions. In: Stability Problems for Stochastic Models, Proceedings, Moscow, VNIISI, 1984, 40 - 54 (in Russian); English transl., J. Soviet Math. 35 (19), 2

[119] Whittle, P.: Optimization Over Time: Dynamic Programming and Stochastic Control. Wiley, New York, 1982

[120] Woyczynski, W. A.: On Marcinkiewicz-Zygmund laws of large numbers in Banach spaces and related rates of convergence. Prob. Math. Statist. 1 (1982), 117 - 131

[121] Zolotarev, V. M.: Approximation of distribution of sums of independent random variables with values in infinite dimensional spaces. Theory Prob. Appl. 21 (1976), 721 - 737

[122] Zolotarev, V. M.: Metric distances in space of random variables and their distributions. Math. USSR Sbornik 30 (1976), 373 - 401

[123] Zolotarev, V. M.: General problems of mathematical models. In: Proceedings of the 41st Session of ISI, New Delhi, 1977, 382 - 401

[124] Zolotarev, V. M.: On pseudomoments. Theory Prob. Appl. 23 (1978), 269 - 278

[125] Zolotarev, V. M.: On properties and relationships between certain types of metrics. Zap. Nauchn. Sem. LOMI 87 (1979), 18 - 35 (in Russian)

[126] Zolotarev, V. M.: Ideal metrics in the problems of probability theory. Austral. J. Statist. 21 (1979), 193 - 208

[127] Zolotarev, V. M.: Probability metrics. Theory Prob. Appl. 28 (1983), 278 - 302

[128] Zolotarev, V. M.: Forward, LNM 982. Springer-Verlag, 1983, pp. VII - XVII

[129] Zolotarev, V. M., and Rachev, S. T.: Rate of convergence in limit theorems for the max-scheme. LNM 1155, pp. 415 - 442. Springer-Verlag, 1985

[130] Zolotarev, V. M.: Contemporary Theory of Summation of Independent Random Variables. Nauka, Moscow, 1986 (in Russian)

S. T. Rachev
Department of Statistics
and Applied Probability
University of California
Santa Barbara, CA 93106
USA

L. Rüschendorf
Institut für Math. Statistik
University of Münster
Einsteinstr. 62
D-4400 Münster