FRÉCHET-BOUNDS AND THEIR APPLICATIONS

Ludger Rüschendorf
Inst. für Math. Statistik
Einsteinstr. 62, D-4400 Münster

Summary. This paper gives a review of Fréchet-bounds and their applications. In section two an approach to the marginal problem and Fréchet-bounds based on duality theory resp. the Hahn-Banach theorem is discussed. Main applications concern the Strassen representation theorem for stochastic orders, the sharpness of the classical Fréchet-bounds, the representation of minimal metrics, couplings of distributions, the Monge-Kantorovic-problem, the construction of random variables with maximum (resp. minimum) sum and variances of the sum, maximally dependent random variables and others. For multivariate marginal systems there is a useful reduction principle and there are some bounds for simple systems, which yield a characterization of the marginal problem for a system of two dimensional marginals in a three-fold product space. In section three we discuss some generalizations of the Young-inequality, which are useful for solving the dual problems of the Fréchet-bounds. A basic notion in this connection is the notion of c-convex functions. As an application one can give a nice characterization of solutions of certain transportation problems. We give a probabilistic proof of some generalizations of the Young- and the Oppenheim-inequality. In section four we discuss some statistical applications and problems. The Huzurbazar conjecture on marginal sufficiency, the problem of the optimal combination of marginal tests and the question of estimation theory in marginal models is considered.

1. Introduction

The marginal model is formally defined as follows. Let $E = E_1 \times \ldots \times E_n$, $\mathfrak{A} = \mathfrak{A}_1 \otimes \ldots \otimes \mathfrak{A}_n$ be a finite product of measure spaces. Let $\mathfrak{E} \subset \mathfrak{P}(\{1,\ldots,n\})$, the system of all subsets of $\{1,\ldots,n\}$ with $\bigcup_{J \in \mathfrak{E}} J = \{1,\ldots,n\}$ and let for $J \in \mathfrak{E}$, $P_J \in M(\prod_{j \in J} E_j)$ be a consistent system of probability measures on $\pi_J(E) = \prod_{j \in J} E_j =: E_J$, where π_J denotes the J-projection from E to E_J. Define the marginal model $M_{\mathfrak{E}}$:

(1) $$M_{\mathfrak{E}} = M(P_J,\ J\in\mathfrak{E}) = \{P\in M\ (E,\mathfrak{A});\ P^{\pi_J} = P_J,\ J\in\mathfrak{E}\}$$

to be the set of all probability measures on E with marginals $P_J = P^{\pi_J}$ of the J-components, $J\in\mathfrak{E}$.

There are some different type of problems of interest in marginal models and related to Frechet-bounds. The <u>marginal problem</u> is the question, whether $M_{\mathfrak{E}} \neq \emptyset$. It was shown by Vorobev (1962), Kellerer (1964), that the property "consistency of $(P_J)_{J\in\mathfrak{E}}$ implies $M_{\mathfrak{E}} \neq \emptyset$" is a purely combinatorial (graphtheoretic) property and is equivalent to the nonexistence of "cycles" in $\mathfrak{E}$. Systems $\mathfrak{E}$ with this property are called <u>decomposable</u> resp. <u>simplicial complex</u> (in [100]). Some related existence problems are investigated in [46], [26], [39], [35]. For non-regular systems the marginal problem is generally not easy to decide except in cases, where explicit constructions are known (cf. [14], [84]). Generally, $M_{\mathfrak{E}}$ is a convex set of probability measures, which in a topological situation with tight P_J is also compact. From a theorem of Douglas (1964), $P \in M_{\mathfrak{E}}$ is an extreme point iff $F = \{\sum_{J\in\mathfrak{E}} f_J \circ \pi_J;\ f_J \in \mathfrak{L}^1(P_J)\}$ is dense in $\mathfrak{L}^1(P)$. For simple marginals more information is known if $n = 2$. Generally, $M_{\mathfrak{E}}$ can be empty, can be a small (even one-point) set or can be a large set of distributions.

In applications $M_{\mathfrak{E}}$ describes a model for systems of n components, where for certain subsystems $J \in \mathfrak{E}$ one knows the distributions P_J "exactly", i.e. there are many joint measurements of these components available. The marginal problem only arises if the specification of P_J is not exact. Of particular relevance for applications is the <u>modelling problem</u>. This means that one should not only solve constructively the marginal problem, but moreover construct submodels $\mathfrak{P} = \{P_\vartheta,\ \vartheta\in\Theta\} \subset M_{\mathfrak{E}}$ with the parameter ϑ specifying interesting aspects of the model like e.g. values of certain dependence measures. An interesting problem in this connection is to find an optimal fit of a probability measure (resp. a density) by an element of $M_{\mathfrak{E}}$ (resp. a corresponding density with fixed marginals). For several distances characterizations of the optimal fit have been derived (cf. [10], [84], [91]), allowing in some cases explicit resp. "approximative" solutions. Measuring the distance by the Kullback-Leibler measure an iterative procedure, the "iterative proportional fitting" (IPF) resp. "scaling projection method" has been investigated in the literature. But so far only in the finite discrete case a valid convergence proof has been found (cf. [10]). Most papers are concerned with the case $\mathfrak{E} = \{\{1\},\dots,\{n\}\}$ of simple marginals. In this case we use the notation

(2) $M_{\mathfrak{C}} = M(P_1,\dots,P_n)$.

Some relevant papers on construction problems are [62], [43], [55], [52], [87], [100].

A third class of problems is to find upper and lower bounds for $\int \varphi\, dP$, $\varphi: E \to \mathbb{R}$ measurable, only based on the knowledge of the marginal structure. The optimal bounds are called <u>Fréchet-bounds</u>, defined by:

(3) $M_{\mathfrak{C}}(\varphi) = \sup\{\int \varphi dP;\ P \in M_{\mathfrak{C}}\},\ m_{\mathfrak{C}}(\varphi) = \inf\{\int \varphi\, dP;\ P \in M_{\mathfrak{C}}\}$.

Since $m_{\mathfrak{C}}(\varphi) = -M_{\mathfrak{C}}(-\varphi)$ it is enough to consider either $M_{\mathfrak{C}}$ or $m_{\mathfrak{C}}$. The classical Fréchet-bounds concern the case of simple marginals and $E_i = \mathbb{R}^1$, $1 \le i \le n$. Then $P \in M(P_1,\dots,P_n)$, if and only if the distribution function $F = F_P$ satisfies:

(4) $\underline{F}(x) \le F(x) \le \overline{F}(x),\ x \in \mathbb{R}$,

where $\underline{F}(x) = (\sum_{i=1} F_i(x_i) - (n-1))_+$, $\overline{F}(x) = \min(F_i(x_i))$. $\underline{F}, \overline{F}$ are the "lower" resp. "upper" Frechet-bounds. $\overline{F}$ is a distribution function (is an element of the Fréchet-class $\mathfrak{F}(P_1,\dots,P_n)$), $\underline{F} \in \mathfrak{F}(P_1,\dots,P_n)$ if $n = 2$, but Dall'Aglio (1972) showed that for $n \ge 3$, $\underline{F}$ is a df only in very exceptional cases. Based on (4) many authors established sharp bounds for $n = 2$ and $\varphi(x,y) = \psi(x-y)$, ψ convex (or concave); in particular $\varphi(x,y) = |x-y|^\alpha$, $\alpha \ge 1$, cf. [30] - [33], [97], [11] - [14], [8], [112], [108], [109], [79], [25]. In particular we refer to the interesting survey article of Dall'Aglio (1972).

More general results on Frechet-bounds can be derived from duality theory. Define the dual problems corresponding to (3)

(5)
$$U(\varphi) := \inf \{\sum_{J\in\mathfrak{C}} \int f_J dP_J;\ \sum_{J\in\mathfrak{C}} f_J \circ \pi_J \ge \varphi\}$$
$$I(\varphi) := \sup \{\sum_{J\in\mathfrak{C}} \int f_J dP_J;\ \sum_{J\in\mathfrak{C}} f_J \circ \pi_J \le \varphi\},$$

then, obviously,

(6) $M_{\mathfrak{C}}(\varphi) \le U(\varphi),\ I(\varphi) \le m_{\mathfrak{C}}(\varphi)$

and the question of equality in (6) and the existence of solutions is interesting. Some general results on this question were derived in [58], [75], [34], [79], [47], [48], [49], [70], yielding explicit results in particular in the case of simple marginals. In the case of multivariate marginals there are only few papers on Fréchet-bounds resp. Fréchet-classes (cf. [14], [111], [96], [98]).

Applications concern almost sure representations of stochastic orders (Strassen's result), construction of maximally dependent random variables, random variables with maximum sums, r.v.'s with minimum variance of the sum (Monte Carlo Simulation), the Monge-Kantorovic mass transportation problem, construction of minimal metrics and optimal couplings and many others. A basic problem for the study of Fréchet bounds is the study of inequalities of the type $\varphi \leq \sum_{J\in\mathfrak{C}} f_J \circ \pi_J$ arising in the definition of the bounds in (5).

We finally mention some statistical problems connected with marginal models. A general question is the following: How can one improve statistical procedures knowing the marginal structure in comparison to the status of ignorance. A different question concerns the robustness of statistical procedures against departures from an ideal independent situation by dependence. There are some close connections between the marginal problem and some recent papers on graphical interaction models, which allow a simplified statistical analysis by their inherent conditional independence properties (cf. [16], [56]). A stochastic ordering result in marginal models allows an easy proof of the Huzurbazar conjecture on partial sufficiency (cf. [95]).

2. Existence and Duality

One method to prove existence and duality results for the marginal problem is to apply some wellknown duality theorems for (topological) vector spaces. This leads to general duality results, where $M_{\mathfrak{C}}(\varphi)$, $m_{\mathfrak{C}}(\varphi)$ are replaced by

(7) $$\tilde{M}_{\mathfrak{C}}(\varphi) = \sup\{\int \varphi dP;\ P \in \widetilde{M_{\mathfrak{C}}}\},$$

where $\tilde{M}_{\mathfrak{C}} := \text{ba}(P_J, J \in \mathfrak{C})$ is the set of finite additive contents with marginals P_J. In a second step one has to establish conditions on φ, resp. the topology, to ensure that

(8) $$\tilde{M}_{\mathfrak{C}}(\varphi) = M_{\mathfrak{C}}(\varphi),\ \tilde{m}_{\mathfrak{C}}(\varphi) = m_{\mathfrak{C}}(\varphi)$$

and to ensure the existence of solutions. This approach has been developed in [75], [76], [79]. The first step can also be based on the Hahn-Banach theorem directly. This has been discussed in greater generality by Lembcke (1972) and Luschgy and Thomsen (1983) (the latter paper also including a discussion on extreme points). The following formulation in Section 2.1 arose from a discussion with H. Luschgy.

2.1. A Generalization of the Marginal Problem

Let on a general measure space $(X,\mathfrak{B})$ (which in this section is not necessarily a product space) $\mathfrak{B}_i \subset \mathfrak{B}$, $i \in I$, be a system of sub-σ-algebras with probability measures $P_i \in M^1(X,\mathfrak{B}_i)$, $i \in I$. Define

$$M = \{P \in M(X, \mathfrak{B});\ P|\mathfrak{B}_i = P_i,\ i \in I\}$$

(9)

$$\tilde{M} = \{P \in ba(X, \mathfrak{B});\ P|\mathfrak{B}_i = P_i,\ i \in I\};$$

$\tilde{M}$ is the set of bounded additive contents with marginals P_i. We assume consistency of (P_i), i.e.

(10) $A \in \mathfrak{B}_{i_1} \cap \mathfrak{B}_{i_2}$ implies that $P_{i_1}(A) = P_{i_2}(A)$.

Furthermore, we define

(11) $F = \{\sum_{i \in I_o} f_i;\ I_o \subset I \text{ finite},\ f_i \in \mathfrak{L}^1(\mathfrak{B}_i, P_i)\} = \bigoplus_{i \in I} \mathfrak{L}^1(\mathfrak{B}_i, P_i)$

the direct sum of the $\mathfrak{B}_i$-measurable functions which are integrable w.r.t. P_i. F is a vector subspace of the vectorspace

(12) $\mathfrak{L}^m = \{\varphi \in \mathfrak{L}(X, \mathfrak{B});\ \exists f \in F \text{ with } \varphi \le f\}$,

the set of measurable functions which are majorized by an element of F. By consistency the linear operator

(13) $T: F \to \mathbb{R}\ ,\ T(\sum_{i \in I_o} f_i) = \sum_{i \in I_o} \int f_i \, dP_i$

is well defined.

Theorem 1. a) (Marginal Problem)
$\tilde{M} \neq \emptyset$ iff $T \ge 0$ (i.e. $f \in F$, $f \ge 0$ implies $Tf \ge 0$).

b) (Duality) For $\varphi \in \mathfrak{L}^m$ we have:

(14) $\tilde{M}(\varphi) := \sup\{\int \varphi \, dP;\ P \in \tilde{M}\} = U(\varphi) := \inf\{Th;\ h \in F,\ \varphi \le h\}$.

c) If $U(\varphi) > -\infty$, then there exists a $P \in \tilde{M}$ with $\tilde{M}(\varphi) = \int \varphi \, dP$.

Proof. a) The direction "⇒" is trivial. For the converse direction observe that U is sublinear on $\mathfrak{L}^m$ and $Uf = Tf$ for $f \in F$. If S is a linear functional on $\mathfrak{L}^m$, $S \le U$, then for $f \in F$, $f \ge 0$, holds: $-Sf = S(-f) \le U(-f) = \inf\{Th;\ -f \le h,\ h \in F\} \le T0 = 0$ i.e. $S \ge 0$ and, obviously, $S|F = T$.

By Hahn-Banach there exists an extension S of T to $\mathfrak{L}^m$, $S \le U$. Riesz' representation theorem ensures the existence of an element $P \in ba(X, \mathfrak{B})$ representing S. Since $S|F = T$, it follows that $P \in \tilde{M}$.

b), c) A corollary to the Hahn-Banach theorem is the existence of an extension S with $S\varphi = U\varphi$ if $U\varphi > -\infty$. The corresponding content then yields b), c) if $U\varphi > -\infty$. If $U\varphi = -\infty$, then also $M(\varphi) = -\infty$; so b) is valid generally.

Remark. Related existence problems are proved similarly. Let e.g. for a finite measure μ $\tilde{M}_\mu := \{P \in ba(X,\mathfrak{B}); P|\mathfrak{B}_i = P_i, i \in I, P \le \mu\}$. Replace the operator from (14) by $U_\mu(\varphi) = \inf \{U(\varphi_o) + \int h_+ d\mu;\ \varphi_o + h \ge \varphi\}$. Then the existence and duality results analogously to (14) are valid (cf. [57]).

Consider next the following assumptions:

A.1 $(X,\mathfrak{B}_i,P_i)$, $i\in I$, are compactly approximable, i.e. there exist compact set-systems $\mathfrak{E}_i \subset \mathfrak{B}_i$ with $P_i(B_i) = \sup \{P_i(E_i);\ E_i \subset B_i,\ E_i \in \mathfrak{E}_i\}$, $i \in I$.

A.2 $(X,\mathfrak{B})$ is a topological space with Borel σ-algebra $\mathfrak{B}$ and $\mathfrak{R} = \mathfrak{R}(\bigcup_{i\in I} \mathfrak{B}_i)$ contains a countable basis of the topology.

Let $\mathfrak{L}^1(X,\mathfrak{R}(\bigcup_{i\in I} \mathfrak{A}_i),P)$ denote the set of P-integrable functions, where P is considered as a content on the algebra $\mathfrak{R}(\bigcup_{i\in I} \mathfrak{B}_i)$ (cf. Dunford, Schwartz (1967), Def. 17, p. 112).

Theorem 2. a) If A.1 holds, then: $M \ne \emptyset$ iff $T \ge 0$. Furthermore, $M(\varphi) = U(\varphi)$ for $\varphi \in \cap \mathfrak{L}^1(X,\mathfrak{R}(\bigcup_{i\in I} \mathfrak{A}_i),P)$.

b) If A.1 and A.2 hold, then $M(\varphi) = U(\varphi)$ for $\varphi \in C_b(X)$.

Proof. A.1 implies that any $P \in \tilde{M}$ is compactly approximable on $\mathfrak{R}(\bigcup_{i\in I} \mathfrak{A}_i)$ and, therefore, σ-additive on $\mathfrak{R}(\bigcup_{i\in I} \mathfrak{A}_i)$, implying the existence of a σ-additive extension. The proof of the duality theorem is similar to [79], [84], Theorem 3. □

Remark. The duality part of Theorem 2 in the case of multivariate marginals was stated in [84] , Theorem 3, for upper and lower semicontinuous functions. The indicated proof is only valid for bounded continuous function. It can presumably be extended to upper semi-continuous functions (one has to prove, that U is σ-continuous for increasing sequences), but the result is not true for lower semicontinuous functions, as was indicated by a counterexample of H. Kellerer. □

2.2. The Case of Simple Marginals. $M_{\mathfrak{E}} = M(P_1,\dots,P_n)$

In the case of simple marginals the duality and existence results of 2.1 have been generalized by Kellerer (1984) to more general functions and spaces. The proofs are based on the study of the continuity properties of $M_{\mathfrak{E}}$, U resp. $m_{\mathfrak{E}}$, I. These continuity properties combined with Choquet's capacity theorem yield in particular the following duality theorem. $(E_i, \mathfrak{A}_i)$ are assumed to be Hausdorff topological spaces with Borel σ-algebras $\mathfrak{A}_i$. This assumption is made throughout the rest of this paper. Also we assume generally that P_J are Radon measures, $J \in \mathfrak{E}$.

Theorem 3. (Kellerer (1984), Theorem 2.21)

The duality theorem $M_{\mathfrak{C}}(\varphi) = U(\varphi)$ is true for

a) $\varphi \in \mathfrak{F}$, the class of upper-semicontinuous functios with values in $\overline{\mathbb{R}}$. In this case there exists a maximal measure,

b) $\varphi \in \mathfrak{L}_m(E, \otimes \mathfrak{A}_i)$, the class of measurable functions w.r.t. $\otimes \mathfrak{A}_i$, which are majorized from below by an element of F (F as in Section 2.1). There exists a solution $f^* \in F$ of the dual problem if $U(\varphi) < \infty$. The product σ-algebra can be replaced by the Baire σ-algebra.

c) $\varphi \in S_m(E)$, the class of lower majorized Suslinfunctions w.r.t. $\mathfrak{F}$. □

Kellerer (1987), Proposition 5.6 also proved a related result for multivariate marginals in the decomposable case.

An interesting consequence of the duality theorem is the following theorem of Strassen (1965) (who gave a proof in the case of polish spaces) saying that for n = 2 one can restrict in the definition of U to two valued functions.

Theorem 4. (Strassen (1963), Kellerer (1984))

Let $n = 2$ and $B \in \mathfrak{A}_1 \otimes \mathfrak{A}_2$, then

(15) $$M_{\mathfrak{C}}(B) = \inf \{P_1(B_1) + P_2(B_2);\ B \subset \bigcup_{i=1}^{2} \pi_i^{-1}(B_i)\}$$

$$m_{\mathfrak{C}}(B) = \sup \{P_1(B_1) + P_2(B_2) - 1;\ B \supset B_1 \times B_2\}.$$

If B is closed, then B_i can be restricted to the class of closed sets. □

We next discuss some more concrete applications of the duality theorem.

2.2.1. Stochastic Orders

Let $n = 2$, $E_1 = E_2 = Y$, $(Y, \leq)$ an ordered topological space with

(16) $$R(Y) = \{(x,y) \in Y \times Y;\ \ x \leq y\}$$

closed, then one obtains from (15) the a.s. representation for the stochastic order $\leq_{s,t}$ w.r.t. monotone increasing functions.

Theorem 5. a) (cf. [47], Prop. 3.11; [85], Lemma 1)

(17) $$M_{\mathfrak{C}}(R(Y)) = 1 - \sup \{P_1(A) - P_2(A);\ A \text{ closed, isotone}\}.$$

b) (Strassen representation theorem, cf. [101])

$P_1 \leq_{st} P_2 \Leftrightarrow \exists P \in M(P_1,P_2)$ with $P(R(Y)) = 1$. □

Strassen's a.s. representation theorem has ben very influential for the theory and applictions of stochastic ordering. It has been extended to ordering results for stochastic processes (cf. [44], [103], [78], it has been extended to "stochastic" ordering not induced by a partial order on Y as e.g. the ordering w.r.t. convex functions (cf. [104] , [77]) and found many applications in the ordering of queues, Markov chains, risk theory and in statistics (cf. [103], [62], [71], [78], [88]).

2.2.2. Sharpness of the Classical Fréchet-Bounds

For product sets $A = A_1 \times \ldots \times A_n \in \mathfrak{A}_1 \otimes \ldots \otimes \mathfrak{A}_n$ there is the obvious generalization of the Fréchet-bounds in (4).

Theorem 6. (cf. [79], Theorem 6)

$$(18)\qquad \begin{aligned} M_{\mathfrak{C}}(A_1 \times \ldots \times A_n) &= \min\{P_i(A_i);\ 1 \leq i \leq n\} \\ m_{\mathfrak{C}}(A_1 \times \ldots \times A_n) &= \Big(\sum_{i=1}^{n} P_i(A_i) - (n-1)\Big)_+ . \end{aligned}$$

Proof. In the case $n = 2$ we obtain from (15)

$$\begin{aligned} M_{\mathfrak{C}}(A_1 \times A_2) &= \inf\{P_1(B_1) + P_2(B_2);\ A_1 \times A_2 \subset B_1 \times E_2 \cup E_1 \times B_2\} \\ &= \min(P_1(A_1),\ P_2(A_2)). \end{aligned}$$

The case $n \geq 2$ is proved by induction. Let $Q_o \in M(P_1,\ldots,P_n)$ satisfy $Q_o(A_1 \times \ldots \times A_n) = \min(P_i(A_i))$, then $\sup\{P(A_1 \times \ldots \times A_{n+1});\ P \in M(P_1,\ldots,P_{n+1})\} = \sup_{Q \in M(P_1,\ldots,P_n)} \sup_{P \in M(Q,P_{n+1})} P(A_1 \times \ldots \times A_{n+1}) \geq \sup_{P \in M(Q_o,P_{n+1})} P(A_1 \times \ldots \times A_{n+1})$

$= \min(Q_o(A_1 \times \ldots \times A_n),\ P_{n+1}(A_{n+1}))$ (from the case $n = 2$)

$= \min(P_i(A_i))$ by induction.

Since the oppositive inequality is trivial, the first part of (18) is proved. The proof of the second part is analogously. □

In [79] the proof was given by direct calculation of the upper resp. lower bounds for the duality theorem. Theorem 6 implies that the Fréchet bounds are identical to the Bonferoni bounds in the following sense. Define $B_i := E_1 \times \ldots \times A_i \times \ldots \times E_n$, $p_i = P_i(A_i)$, then $A_1 \times \ldots \times A_n = \bigcap_{i=1}^{n} B_i$. The Bonferoni bounds are defined by

(19) $$\tilde{U}((p_i)) = \sup\{P(\bigcap_{i=1}^{n} B_i);\ B_i \in \mathfrak{A},\ P(B_i) = p_i,\ 1 \le i \le n\}$$
$$\tilde{L}((p_i)) = \inf\{P(\bigcap_{i=1}^{n} B_i);\ B_i \in \mathfrak{A},\ P(B_i) = p_i,\ 1 \le i \le n\}.$$

So the knowledge of the whole marginal distribution does not help to obtain better bounds for product sets in comparision to knowing only the probabilities $P_i(A_i)$. For the ordering by survival functions

(20) $$P \le_s Q \quad \text{if} \quad P([x,\infty)) \le Q([x,\infty))$$

for all $x \in \mathbb{R}^n$ it has been proved that

(21) $$P \le_s Q \quad \text{iff} \quad \int \varphi\, dP \le \int \varphi\, dQ$$

for Δ-monotone (in pairs) resp. quasimonotone resp. L-superadditive functions (cf. Cambanis, Simons and Stout (1976) and Whitt (1976) for n = 2, Rüschendorf (1979, 1981, 1983), Tchen (1980), Marshall, Olkin (1979), Mosler (1982) for $n \ge 2$), (21) combined with (18), (4) imply sharp results for $M_{\mathfrak{C}}(\varphi)$. These are related to rearrangement inequalities (cf. [112], [81]). The case $n = 2$, $\varphi(x,y) = -\psi(x-y)$, ψ convex, is due to Bertino (1966). Some partial results are in [79] for the lower bound $m_{\mathfrak{C}}(\varphi)$. An open problem is e.g. to determine $m_{\mathfrak{C}_3}(\varphi)$ for $Y = [0,1]$, $P_i = R(0,1)$, the uniform distribution $1 \le i \le 3$ and $\varphi(x) = \prod_{i=1}^{3} x_i$. □

2.2.3. Representation of Minimal Metrics

Some of the wellknown probability metrics have a representation as a "minimal metric".

2.2.3.1. Levy-Prohorov-Metric

On a metric space (Y,d) with Borel σ-algebra $\mathfrak{B}$ define for $A \in \mathfrak{B}$, $\varepsilon > 0$,

(22) $$A^{\varepsilon} := \{y \in Y;\ d(x,y) < \varepsilon, \ \text{for some} \ x \in A\},\ A^{0} := \bar{A}.$$

From (15) we obtain

Theorem 7. (cf. Dudley (1976), Theorem 18.2)

Let $P_1, P_2 \in M^1(Y,\mathfrak{B})$, $\varepsilon \ge 0$;

a) $\delta > 0$. There exists $P \in M(P_1,P_2)$ with $P\{(x,y) \in Y \times Y;\ d(x,y) > \varepsilon\} < \delta$
$\Leftrightarrow \forall A \in \mathfrak{F} = \mathfrak{F}(Y)$: $P_1(A) \le P_2(A^{\varepsilon}) + \delta$;

b) $\delta \ge 0$. There exists $P \in M(P_1,P_2)$ with $P\{d(x,y) > \varepsilon\} \le \delta$
$\Leftrightarrow P_1(A) \le P_2(A^{\varepsilon}) + \delta,\ \forall A \in \mathfrak{F}(Y)$. □

Theorem 7 implies the Strassen-representation of the Levy-Prohorov metric

(23) $\pi(P_1,P_2) = \inf\{\varepsilon > 0:\ P_1(A) \le P_2(A^{\varepsilon}) + \varepsilon,\ \forall A \in \mathfrak{B}(Y)\}$.

Define for $P \in M^1(Y \times Y, \mathfrak{B} \otimes \mathfrak{B})$ the Ky-Fan (probability-) metric

(24) $K(P) = \inf\{\varepsilon > 0:\ P(d(x,y) > \varepsilon) < \varepsilon\}$

and consider the corresponding minimal metric

(25) $\hat{K}(P_1,P_2) = \inf\{K(P);\ P \in M(P_1,P_2)\}$.

Theorem 8. (Strassen (1964), Dudley (1968))

(26) $\hat{K} = \pi$. □

π metrizes the topology of weak convergence on the set of tight Borel measures (this is immediate from Dudley (1976), Theorem 8.3, who considers the case of separable metric spaces), i.e.

(27) $P_n \xrightarrow{\mathfrak{D}} P$ if and only if $\pi(P_n,P) \to 0$.

A basic coupling result is the almost sure representation theorem. The proof of part b) makes essential use of Theorem 8.

Theorem 9. (Almost sure representation theorem)

a) (Skorohod, Strassen, Dudley, Wichura, cf. [23])

Let $P_n, P \in M^1(Y,d)$ be tight. Then $\pi(P_n,P) \to 0$ if and only if there exists a probability space $(\Omega, \mathfrak{A}, R)$ and Y-valued random variables X_n, X on $(\Omega, \mathfrak{A})$, such that $R^{X_n} = P_n$, $R^X = P$ and $d(X_n,X) \to 0$ a.s.

b) Rachev, Rüschendorf and Schief (1988), Dudley (1989)

If $P_n, Q_n \in M^1(Y,d)$ are tight and $\pi(P_n,Q_n) \to 0$, then $d(X_n,Y_n) \to 0$ a.s. for some versions X_n, Y_n on a probability space $(\Omega, \mathfrak{A}, R)$ with $R^{X_n} = P_n$, $R^{Y_n} = Q_n$. □

Remark. If $P_n \in M(Y)$, $Y = Y_1 \times Y_2$ a product space, $P_n \in M(Q_n,R)$ and $P_n \xrightarrow{\mathfrak{D}} P$, then the following sharpening of Theorem 8 is not true: "There exist versions of P_n of the form (X_n,Z), such that (X_n) is a.s. convergent (cf. [70])." □

2.2.3.2. $\mathfrak{L}^p$-Metrics

Define the probability metric

(28) $\mathfrak{L}^{\infty}(P) = \operatorname{ess\,sup}_P d(x,y) = \inf\{\varepsilon > 0;\ P(d(x,y) > \varepsilon) = 0\}.$

Theorem 7.b) implies the following representation of the corresponding minimal metric.

Theorem 10. (cf. Dudley (1976), Theorem 18.2)

(29) $$\widehat{\mathfrak{L}^{\infty}}(P_1,P_2) := \inf\{\mathfrak{L}^{\infty}(P);\ P \in M(P_1,P_2)\} = \inf\{\varepsilon > 0;\ P_1(A) \le P_2(A^{\varepsilon}),\ \forall A \in \mathfrak{B}(Y)\}.$$ □

For the $\mathfrak{L}^p$-distance, $1 \le p < \infty$

(30) $$\mathfrak{L}^p(P) = \int d^p(x,y)\,dP(x,y)$$

(the corresponding probability metric is $d_p(P) = (\mathfrak{L}^p(P))^{1/p}$) the duality theorem of Section 2.2 implies for $P_1,P_2 \in M^1(Y)$ with $\int d^p(x,a)dP_i(x) < \infty$

(31) $$\widehat{\mathfrak{L}^p}(P_1,P_2) = \sup\{\int f\,dP_1 + \int g\,dP_2;\ f \in \mathfrak{L}(P_1),\ g \in \mathfrak{L}(P_2),\ f(x) + g(y) \le d^p(x,y)\}$$

(cf. also [68]). For $p = 1$ there is the following strengthening of (31).

Theorem 11. (Kantorovic-Rubinstein-Theorem, cf. [46], [58], [106], [107], [28], [47], [70]).

If $\int d(x,a)(P_1 + P_2)(dx) < \infty$, then

(32) $$\widehat{\mathfrak{L}^1}(P_1,P_2) = \sup\{\int f\,d(P_1 - P_2);\ f \in \mathrm{Lip}(Y)\},\quad \text{where}$$ $$\mathrm{Lip}(Y) = \{f: Y \to \mathbb{R}^1;\ |f(x) - f(y)| \le d(x,y)\}.$$

Proof. From (31)

$$\hat{\mathfrak{L}}^1(P_1,P_2) = \sup\{\int f_1\,dP_1 - \int f_2\,dP_2;\ f_1(x) - f_2(y) \le d(x,y),\ f_i \in \mathfrak{L}^1(P_i)\}.$$

Let $P_i = R + R_i$, $i = 1,2$, be a decomposition of P_i, $i = 1,2$, where the measures R_i are orthogonal with supports A_1, A_2. If $R = 0$, then define

$$f(x) := \begin{cases} \sup\{f_1(x_1) - d(x_1,x);\ x_1 \in A_1\} & \text{if } x \in A_2 \\ \inf\{f_2(x_2) + d(x_2,x);\ x_2 \in A_2\} & \text{if } x \in A_1 \end{cases}$$

Then $f \in \mathrm{Lip}(Y)$ and f is better then f_1, f_2, i.e. $f(x) \ge f_1(x)$, $x \in A_1$, $f(x) \le f_2(x)$, $x \in A_2$, and, therefore, $\int f_1 dP_1 - \int f_2 dP_2 = \int_{A_1} f_1 dR_1 - \int_{A_2} f_2 dR_2 \le \int f\,d(R_1 - R_2) = \int f\,d(P_1 - P_2)$ and from (31) $\widehat{\mathfrak{L}^1}(P_1,P_2) = l(P_1,P_2) := \sup\{\int f\,d(P_1 - P_2);\ f \in \mathrm{Lip}(Y)\}$. If $R \ne 0$, $R_i \ne 0$, then the result follows from the following relations: $\widehat{\mathfrak{L}^1}(P_1,P_2) \ge l(P_1,P_2) = l(R_1,R_2) = \widehat{\mathfrak{L}^1}(R_1,R_2) := \inf\{\int d(x,y)d\tilde{R}(x,y),\ \tilde{R} \in M(R_1,R_2)\} \ge \inf\{\int d(x,y)dP(x,y);\ P \in M(P_1,P_2)\} = \widehat{\mathfrak{L}^1}(P_1,P_2)$. For the last inequality observe that for $\tilde{R} \in M(R_1,R_2)$ and $Q \in M(R,R)$ concentrated on the diagonal, $P := Q + \tilde{R} \in M(P_1,P_2)$ and $\int d(x,y)dP = \int d(x,y)d\tilde{R}$. If $R_i = 0$, then $P_1 = P_2$ and the equality is trivial. □

The idea of this proof is due to Szulga (1978). The integrability assumption $\int d(x,a)(P_1 + P_2)(dx) < \infty$ was removed by Kellerer (1984). For $Y = \mathbb{R}^1$ the minimal ℓ^p metrics are explicitely known (cf. Gini [36], Salvemini [97], Dall'Aglio [11] - [14], Fréchet [30] - [33], Hoeffding [40], Vallender [109]).

$$(33) \qquad \widehat{\ell^1}(P_1,P_2) = \int |F_1(x) - F_2(x)| dx$$

$$\widehat{\ell^p}(P_1,P_2) = \int_0^1 |F_1^{-1}(t) - F_2^{-1}(t)|^p dt, \ p \geq 1,$$

where F_i are the df's of P_i.

For $Y = \mathbb{R}^k$ there are few explicit solutions. Let $|x|$ denote the euclidean norm in $\mathbb{R}^k$.

<u>Theorem</u> 12. a) (Knott and Smith (1984), Rüschendorf and Rachev (1990)) If $\int |x|^2 dP_i(x) < \infty$, $i = 1,2$, then random variables X, Y with distributions P_1, P_2 satisfy: $E|X - Y|^2 = \widehat{\ell^2}(P_1,P_2) \Leftrightarrow \exists\, f: \mathbb{R}^k \to \mathbb{R}^1$ closed, convex such that a.s. $Y \in \partial f(X)$, the subgradient of f in X.

b) (Dawson and Landau (1982), Olkin and Pukelsheim (1982), Givens and Shortt (1984))

For nonsingular covariance matrices Σ_1, Σ_2 and $a_1, a_2 \in \mathbb{R}^k$:

$$(34) \qquad \widehat{\ell^2}(N(a_1,\Sigma_1), N(a_2,\Sigma_2)) = |a_1 - a_2|^2 + \operatorname{tr}\Sigma_1 + \operatorname{tr}\Sigma_2 - 2\operatorname{tr}(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2})^{1/2}.$$

□

<u>Remark.</u>

a) A differentiable continuous function f is convex, if and only if

$$(35) \qquad \Phi(x) := \nabla f(x) \text{ is monotone, i.e. } \langle x - y, \Phi(x) - \Phi(y)\rangle \geq 0, \ \forall x,y$$

(cf. [74], p. 99). Therefore, if X has distribution P_1, if Φ is the gradient of a differentiable function f, $\Phi(X) = \nabla f(X)$, has distribution P_2, then $(X,\Phi(X))$ is an optimal coupling w.r.t. ℓ^2-distance if and only if Φ is monotone. Let $\omega = \sum_{i=1}^k \Phi_i dx_i$, then $d\omega = \sum_{i<j} \left(\frac{\partial \Phi_j}{\partial x_i} - \frac{\partial \Phi_i}{\partial x_j}\right) dx_i \wedge x_j$ and by Poincaré's lemma we obtain: If Φ is continuously differentiable, then $(X,\Phi(X))$ is an optimal coupling w.r.t. ℓ^2-distance, if and only if

(36) 1. $\frac{\partial \Phi_i}{\partial x_i} = \frac{\partial \Phi_i}{\partial x_j}$, $\forall i \neq j$ and

2. Φ is monotone.

For linear functions $\Phi(x) = Ax$, this is equivalent to the assumption that A is positive semidefinite and symmetric. In the normal case (34) with $a_1 = a_2 = 0$, we obtain with $\Phi(x) = \Sigma_1^{-1/2}\Sigma_2^{1/2}x$: $(X,\Phi(X))$ is optimal if $\Sigma_1\Sigma_2 = \Sigma_2\Sigma_1$. In the general case we use $\Phi(x) = \Sigma_1^{1/2}(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2})^{-1/2}\Sigma_1^{1/2}x$ to obtain (34).

b) The proof of Theorem 12. a) can be based on the duality theorem and results from convex analysis (for some extensions cf. Section 3). The minimal $\mathfrak{L}^p$-metrics are special instances of the Monge-Kantorovic mass transference problem. A review of this type of problems and several further representations of minimal metrics are given in Rachev (1985). Some results for the $\mathfrak{L}^2$-metrics and their application to approximation problems are discussed in [86].

2.2.4. Probability on Diagonals

For $n \geq 2$, $E_1 = \dots = E_n = Y$ and for $A \in \mathfrak{A}_1$ let

(37) $$\Delta_n(A) := \{(x,\dots,x);\ x \in A\} \in \bigotimes_{i=1}^n \mathfrak{A}_1$$

and let

(38) $$P_1 \wedge \dots \wedge P_n(A) := \inf\ \{\sum_{i=1}^n P_i(A_i);\ A_i \in \mathfrak{A}_1,\ \sum_{i=1}^n A_i = A\}.$$

Theorem 13. $\exists P^* \in M(P_1,\dots,P_n)$ such that for all $A \in \mathfrak{A}_1$:

(39) $$P^*(\Delta_n(A)) = M_{\phi}(\Delta_n(A)) = P_1 \wedge \dots \wedge P_n(A).$$

Proof. For n = 2 the equality $M_{\phi}(\Delta_2(A)) = P_1 \wedge P_2(A)$ follows from (15). Since $P_1 \wedge P_2$ is a measure on Y, M_{ϕ} is additive on $\mathfrak{A}_1$. This implies by an inductive argument the existence of $P^* \in M(P_1,P_2)$ with $P^*(\Delta_2(A)) = P_1 \wedge P_2(A)$, $\forall A \in \mathfrak{A}_1$ (If $P^*(A_1 + A_2) = M_{\phi}(A_1 + A_2) = M_{\phi}(A_1) + M_{\phi}(A_2)$, then $P^*(A_i) = M_{\phi}(A_i)$, i = 1,2).

If $n \geq 2$, then take $Q \in M(P_1,\dots,P_{n-1})$ with $Q(\Delta_{n-1}(A)) = P_1 \wedge \dots \wedge P_{n-1}(A)$, $A \in \mathfrak{A}_1$, (induction hypothesis). Then $M_{\phi}(\Delta_n(A)) = \sup_{Q' \in M(P_1,\dots,P_{n-1})} \sup_{P \in M(Q',P_n)} P(\Delta_n(A))$

$\geq \sup_{P \in M(Q,P_n)} P(\Delta_n(A)) = \inf\{Q(B_1) + P_n(B_2);\ \Delta_n(A) \subset B_1 \times Y \cup Y \quad \times B_2\}$

$= \inf\{Q(\Delta_n(A_1)) + P_n(A \setminus A_1);\ A_1 \subset A\} = \inf\{P_1 \wedge \dots \wedge P_{n-1}(A_1) + P_n(A \setminus A_1);\ A_1 \subset A\}$
$= P_1 \wedge \dots \wedge P_n(A)$. The oppositive inequality is trivial. □

In the case $n=2$, $A=Y$, (39) implies with $\Delta_2 = \Delta_2(Y)$ and

(40) $$d_v(P_1,P_2) = \sup\{P_1(B) - P_2(B);\ B \in \mathfrak{A}_1\},$$

the wellknown representation of the sup-metric d_v:

(41) $$d_v(P_1,P_2) = 1 - M_{\mathfrak{C}}(\Delta_2) = m_{\mathfrak{C}}(\Delta_2)$$

due to Dobrushin (1969).

2.2.5. Random Variables With Maximum Sums

Problem: For $P_i \in M(\mathbb{R})$ with df's F_i, $1 \le i \le n$, determine the maximum resp. minimum probability of

(42) $$A_n(t) := \{x \in \mathbb{R}^n;\ \sum_{i=1}^n x_i \le t\}.$$

This problem was solved independently by Makarov (1981) and Rüschendorf (1982) for n = 2. (For a different proof cf. also Frank, Nelsen and Schweizer (1987).) Makarov introduces this problem as "Kolmogorov's problem".

Theorem 14. (Makarov (1981), Rüschendorf (1982))
For $n=2$ and $t \in \mathbb{R}$ we have

(43) $$M_{\mathfrak{C}}(A_2(t)) = F_1 \wedge F_2(t) = \inf_x (F_1(x-) + F_2(t-x))$$ the infimal convolution.

(44) $$m_{\mathfrak{C}}(A_2(t)) = F_1 \vee F_2(t-) - 1, \text{ where } F_1 \vee F_2(t)) = \sup_x (F_1(x-) + F_2(t-x)).$$

□

For $n \ge 2$ there are some particular results in [77], obtained by explicit solution of the dual problem. If e.g. $P_1 = \dots = P_n = R(0,1)$, then

(45) $$M_{\mathfrak{C}}(A_n(t)) = \frac{2}{n} t,\ 0 \le t \le \frac{n}{2},$$

(46) $$m_{\mathfrak{C}}(A_n(t)) = \min\{(\tfrac{2}{n} t - 1)_+, 1\},\ t \ge 0.$$

If $P_i = \mathfrak{B}(1,\vartheta)$, $1 \le i \le n$, then

(47) $$M_{\mathfrak{C}}(A_n(k)) = \frac{n}{n-k}(1-\vartheta),\ k \le n\vartheta,$$

the solution $P^* \in M_{\mathfrak{C}}$ being a mixture of the uniform distribution on $\{x: \sum_{i=1}^n x_i = k\}$ and a one point measure in $(1,\dots,1)$.

Similar formulas are possible for other geometric objects like circles or triangles (for n = 2).

2.2.6. Monte-Carlo-Simulation

Problem: For $P_i \in M^1(\mathbb{R}^1)$ construct rv's $X_i^* \sim P_i$ with

(48) $$\mathrm{Var}\,(\sum_{i=1}^n X_i^*) \le \mathrm{Var}\,(\sum_{i=1} X_i) \text{ for any } X_i \sim P_i.$$

For n = 2 a solution is the wellknown method of "antithetic variates" (cf. Hammersley Handscomb (1964)). For $n \geq 2$ there are some particular results.

1. If $P_i = R(0,1)$, then it is possible to construct $X_i^* \sim P_i$, $1 \leq i \leq n$, with $\sum_{i=1}^n X_i^* = \frac{n}{2}$ (cf. Gaffke and Rüschendorf (1981)). So (X_i) solve (48) trivially.

2. If P_i are uniform on $\{1,\dots,n\}$, then one can construct X_i^*, $1 \leq i \leq n$, with $\sum_{i=1}^n X_i^* \in \{a, a+1\}$, which solve (48) (cf. [82]).

3. If $P_i = \mathfrak{B}(1,\vartheta)$, then one can again construct $X_i^* \sim P_i$ with $\sum_{i=1}^n X_i^* \in \{k, k+1\}$, $\frac{k}{n} \leq \vartheta \leq \frac{k+1}{n}$, which solve (48). The minimal value of the variance equals the cyclic function

(49) $$v_k(\vartheta) = a(k,\vartheta)(1 - a(k,\vartheta)),\quad a(k,\vartheta) := k\vartheta \pmod 1$$

(cf. Snijders (1984)).

In these examples it is possible to concentrate the distribution of ΣX_i^* "close" to nEX_1^*. For a symmetric distribution (like $N(a,\sigma^2)$) and $n = 2m$ one can choose rv's X_i with $\sum_{i=1}^n X_i = nEX_1$.

For $P_i \in M^1(\mathbb{R}^k, \mathfrak{B}^k)$, $1 \leq i \leq n$, we can similarly consider $\frac{1}{n} \sum_{i=1}^n X_i$ as simulation for $a = \frac{1}{n} \sum_{i=1}^n EX_i$ (typically: $P_1 = \dots = P_n$) with error $E|\frac{1}{n} \Sigma X_i - a|^2$. The corresponding problem is to determine the minimum of $\sum_{i<j} E\langle X_i, X_j\rangle$. For $n = 2$ we obtain a characterization of a solution from Theorem 12 (cf. also Section 3.1): $E\langle X_1^*, X_2^*\rangle = \min_{X_1 \sim P_1, X_2 \sim P_2} E\langle X_1, X_2\rangle$
$\Leftrightarrow \exists f: \mathbb{R} \to \mathbb{R}$ closed, convex, such that $X_2^* \in \partial f(-X_1^*)$.

2.2.7. Maximally Dependent Random Variables

Lai and Robbins (1978) constructed for given $P_i \in M^1(\mathbb{R}^1)$, $i \in \mathbb{N}$, random variables $X_i^* \sim P_i$ such that

(50) $$P^{\max_{1\leq i\leq n} X_i} \leq_{st} P^{\max_{1\leq i\leq n} X_i^*}, \quad \forall n \in \mathbb{N},$$

$X^* = (X_i)$ is called maximally dependent sequence. In the case $P_i = R(0,1)$ there is a nice geometric construction (cf. also [76]). In terms of limit theorems Lai and Robbins established that $\max_{1\leq i\leq n} X_i^*$ is not much larger than $\max \tilde{X}_i$, where $(\tilde{X}_i)$ is an independent sequence (in the case $P_i = P_1$, $\forall i$). For a construction based on duality theory cf. [34] , [50]. From (18) one obtains $P(\max_{1\le i\le n} X_i^* \le t) = (\sum_{i=1}^n F_{P_i}(t) - (n-1))_+$. Solutions then can be defined iteratively.

In the case n = 2, $P_i = R(0,1)$, i = 1,2, $M(P_1,P_2)$ is called the class of doubly stochastic measures. Let U be a R(0,1)-distributed random variable and for a λ^1-preserving transformation g: [0,1] → [0,1] define P_g to be the distribution of (U,g(U)), ${}_gP$ to be the distribution of (g(U),U). If g is one to one then ${}_gP, P_g$ are called permutation measures since $P_g(A \times B) = \lambda^1(A \cap g^{-1}(B))$, $A,B \in \mathfrak{B}^1[0,1]$ and ${}_gP = P_{g^{-1}}$.

The only monotonic transformations of [0,1] which are λ^1 preserving are $g_1(u) = u[\lambda^1]$, $g_2(u) = 1 - u[\lambda^1]$, the corresponding permutation measures are the Fréchet-distributions. The property of two random variables X, Y that Y = g(X), g λ^1-preserving was introduced by Lancaster (1963) under the notation: Y is <u>completely dependent</u> on X. By Theorem 1 of Brown (1966), $M(P_1,P_2)$ is the closure of the set of all permutation measures w. r.t. weak operator topology on L^1, i.e. w.r.t. convergence of integrals of functions $f(x)\cdot g(y) \in L^1(\lambda^2)$. This theorem implies in particular that each doubly stochastic measure (also the product measure) can be approximated w.r.t. convergence in distribution by a sequence of permutation measures and it is easy to give an explicit constructon of an approximation sequence (cf. also Kimeldorf and Sampson (1978)). So in a certain sense complete dependence is close to independence. This is related to the generation of chaotic (stochastic) behaviour of dynamical systems by deterministic models.

2.3. The Case of Multivariate Marginals

In the case of multivariate marginals there are few explicit results. In the decomposable case there is an interesting reduction principle which is proved in [96] for Borel spaces (the proof being valid for universally measurable separable metric spaces). Let $h_i : E_i \to W_i$ be measurable, E_i, W_i Borel spaces, $1 \le i \le n$, let $h_J := (h_j)_{j\in J}: \prod_{j\in J} E_j \to \prod_{j\in J} W_j$, $J \subset \{1,\dots,n\}$, $h = (h_1,\dots,h_n)$.

<u>Theorem</u> 15. (cf. [93])

If $\mathfrak{E}$ is decomposable, then

$$M_{\mathfrak{E}}^h = \{P^h;\ P \in M_{\mathfrak{E}}\} = M(P_J^{h_J};\ J \in \mathfrak{E}). \qquad \square \tag{51}$$

For the special case $M(P_1,\dots,P_n)^h = M(P_1^{h_1},\dots,P_n^{h_n})$ cf. also Rachev and Rüschendorf (1986), Scarsini (1989). If $h_i: ([0,1],\lambda) \to (E_i,\mathfrak{A}_i,P_i)$ with $(\lambda^1)^{h_i} = P_i$, then any $P \in M(P_1,\dots,P_n)$ has a representation $P = Q^h$, $Q \in M(R(0,1),\dots,R(0,1))$. If $E_i = \mathbb{R}$, $h_i(x_i) = F_i^{-1}(x_i)$, $x_i \in (0,1)$, where F_i are the df's of P_i, then $\lambda^{h_i} = P_i$ and $P = Q^h$. Therefore,

$$F_P(x) = Q(h \le x) = Q(F_i^{-1} \le x_i,\ 1 \le i \le n) = F_Q(F_1(x_1),\dots,F_n(x_n)), \tag{52}$$

F_Q is the so called "copula".

(51) implies in particular for the case of simple marginals and $h_i: E_i \to \mathbb{R}^1$, $h_i \in \mathfrak{L}^1(P_i)$:

$$M_{\mathfrak{C}}\Big(\prod_{i=1}^n h_i\Big) = \int_0^1 \prod_{i=1}^n F_{h_i}^{-1}(u)\,du, \tag{53}$$

where F_{h_i} is the df. of $P_i^{h_i}$.

For some decomposable cases in [14], [96] sharp bounds have been proved as e.g. for star-configurations $\mathfrak{C}_1 = \{\{1,j\},\ 2 \le j \le n\}$ or simple series-configurations $\mathfrak{C} = \{\{1,2\},\{2,3\}\}$. In [96] is a discussion of two principles of deriving bounds, the method of Bonferoni-type bounds and the method of conditioning.

In the nonregular case the set $M_{\mathfrak{C}}$ can be empty, can contain one element (uniqueness) or can be a large convex set. This is in contrast with the decomposable case. A further difference is the fact that the continuity properties of $M_{\mathfrak{C}}$, U in the nonregular case seem to be strictly weaker than in the regular case. But these properties need a more detailed investigation.

<u>Example.</u> Let $n = 3$, $\mathfrak{C} = \{\{1,2\},\{2,3\},\{1,3\}\}$, the simplest nonregular case, and let $E_i = [0,1]$. If $P_{ij} = P^{(U,1-U)}$ for all i,j, where $P^U = R(0,1)$ is uniform on (0,1), then $M_{\mathfrak{C}} = \emptyset$.

If $P \in M([0,1]^3)$ with marginals (P_{ij}), $i,j \le 3$, and $P\{x: \Sigma x_i = c\} = 1$, then $M_{\mathfrak{C}} = \{P\}$. For the proof note that for any $Q \in M_{\mathfrak{C}}$ we have $\int (\sum_{j=1}^3 x_j - c)^2\,dQ(x) = \int (\Sigma x_j - c)^2\,dP(x) = 0$ i.e. $Q\{x: \Sigma x_j = c\} = 1$. This implies that the conditional distributions $P^{\pi_3 | \pi_1 = x_1, \pi_2 = x_2} = Q^{\pi_3 | \pi_1 = x_1, \pi_2 = x_2}$ and, therefore, $P = Q$.

If $P_{ij} = R(0,1) \otimes R(0,1)$ for all i,j, then $\lambda^3_{[0,1]} = R(0,1) \otimes R(0,1) \otimes R(0,1) \in M_{\mathfrak{C}}$. Let $v_i: [0,1] \to [-1,1]$ satisfy $\int v_i(x)\,dx = 0$, $1 \le i \le 3$. The measures $P_v := (1 + \prod v_i(x))\lambda^3_{[0,1]}$, $v = (v_i)$, all have two dimensional marginals $\lambda^2_{[0,1]}$, i.e. $P_v \in M_{\mathfrak{C}}$. One can explicitely construct all elements of $M_{\mathfrak{C}}$ which are continuous w.r.t. $\lambda^3_{[0,1]}$ (cf. [87]). □

For $P_{ij} \in M^1(E_i \times E_j)$ let $P_{i|x_j} = P_{ij}^{\pi_i|\pi_j = x_j}$ be the conditional distribution and define for $A \in \mathfrak{A}_1 \otimes \mathfrak{A}_3$

(54) $$U_{13|x_2}(A) := \inf \{P_{1|x_2}(A_1) + P_{3|x_2}(A_3);\ A \subset A_1 \times E_3 \cup E_1 \times A_2\}$$

$$L_{13|x_2}(A) := \sup \{P_{1|x_2}(A_1) + P_{3|x_2}(A_3);\ A \supset A_1 \times A_3\}.$$

Theorem 16. (cf. [96])

a) If $\mathfrak{E} = \{\{1,2\}, \{2,3\}\}$, $B \in \mathfrak{B}_1 \otimes \mathfrak{B}_2 \otimes \mathfrak{B}_3$, then

(55) $$M_{\mathfrak{E}}(B) = \int U_{13|x_2}(B_{x_2}) dP_2(x_2)$$

$$m_{\mathfrak{E}}(B) = \int L_{13|x_2}(B_{x_2}) dP_2(x_2).$$

b) If $\mathfrak{E} = \{\{1,2\}, \{1,3\}, \{2,3\}\}$, then

(56) $$M_{\mathfrak{E}}(\varphi) \le U(\varphi) := \min \{ \int U_{23|x_1}(\varphi_{x_1}) dP_1(x_1),\ \int U_{13,x_2}(\varphi_{x_2}) dP_2(x_2),\ \int U_{12|x_3}(\varphi_{x_2}) dP_3(x_3)\}$$

for $\varphi \in \mathfrak{L}_m(E)$, where $U_{ij|x_k}(\varphi_{x_k})$ are defined analogously to (54). □

From (56) for $\varphi = 1_A$, $A = A_1 \times A_2 \times A_3$ follows

(57) $$M_{\mathfrak{E}}(A) \le \tilde{U}(1_A) \le \min (P_{ij}(A_i \times A_j)),$$

the right hand side being the Bonferoni bound. The last inequality typically is strict. This is in contrast to the case of simple marginals.

In the case $\mathfrak{E} = J_2^3 := \{\{1,2\}, \{1,3\}, \{2,3\}\}$ let

(58) $$C(P_{12},P_{23}) = \{P_{13};\ M(P_{12},P_{13},P_{23}) \neq \emptyset\}$$

be the compatibility set of P_{12},P_{23}. Dall'Aglio (1959, 1972) proved that in the case $E_i = \mathbb{R}^1$:

(59) $$F_{13}(x_1,x_3) := \int \max (F_{1|x_2}(x_1) + F_{3|x_2}(x_3) - 1,0) dP_2(x_2)$$

$$\le F_{13}(x_1,x_3) \le F_{13}(x_1,x_3) = \int \min (F_{1|x_2}(x_3), F_{3|x_2}(x_3)) dF_2(x_2),$$

F_{13}, F_{13} are the minimal and maximal df's of elements of $C(P_{12},P_{23})$. For the converse Dall'Aglio (1959) gives a counterexample.

The following result gives a characterization of the marginal problem $\mathfrak{E} = I_2^3$.

Theorem 17. (cf. [96])

$P_{13} \in C(P_{12},P_{23}) \Leftrightarrow \forall \varphi = \varphi(x_1,x_3)$ bounded, measurable:

(60) $$\tilde{L}_{13}(\varphi) \le \int \varphi \, dP_{13} \le \tilde{U}_{13}(\varphi), \text{ where}$$
$$\tilde{U}_{13}(\varphi) = \int U_{13|x_2}(\varphi) dP_2(x_2), \quad \tilde{L}_{13}(\varphi) = \int L_{13|x_2}(\varphi) dP_2(x_2). \qquad \square$$

It is not sufficient to consider indicator functions only.

3. Inequalities of the Type: $c(x,y) \le f(x) + g(y)$

In this section motivated by the duality theorem we investigate generalizations of the Young inequality (cf. Section 3.2 for a statement).

3.1. c-Convex Functions

For $n = 2$, $\mathfrak{E} = \{\{1\}, \{2\}\}$, $P_1, P_2 \in M(Y)$ and $c = c(x,y) \in \mathfrak{L}_m$ from Theorem 3 follows:

(61) $$M_{\mathfrak{E}}(c) = \inf\{\int f \, dP_1 + \int g \, dP_2;\ f \in \mathfrak{L}^1(P_1),\ g \in \mathfrak{L}^1(P_2),\ c(x,y) \le f(x) + g(y)\}$$

and there exist solutions of the dual problem, if $\int c(x,a)(P_1 + P_2)(dx) < \infty$. A "maximal" measure $P \in M_{\mathfrak{E}}$ exists if $c \in \mathfrak{F}(Y \times Y)$ i.e. c is upper semi-continuous.

If (f,g) are admissible (i.e. $f(x) + g(y) \ge c(x,y)$) and $P \in M_{\mathfrak{E}}$, then $(P,(f,g))$ are solutions if and only if

(62) $$c(x,y) = f(x) + g(y) \quad [P].$$

Therefore, for the calculation of the Frechet-bounds one needs sharp inequalities of the type $c(x,y) \le f(x) + g(y)$.

For $c(x,y) = \pm \langle x,y \rangle$, $x,y \in \mathbb{R}^k$, a theory of these inequalities has been established in the convex conjugate duality theory (cf. Rockafellar (1970)). This led in Theorem 12 to a characterization of optimal couplings w.r.t. $\mathfrak{L}^2$-distance. For general $c: E_1 \times E_2 \to \mathbb{R}^1$, there are several papers, but the results are less complete. For the literature we refer to [19], [27], [42], [1].

For $f: E_1 \to \mathbb{R}^1$ define the c-conjugate

(63) $$f^*: E_2 \to \overline{\mathbb{R}}^1, \quad f^*(y) = \sup_{x \in E_1} (c(x,y) - f(x))$$

and the doubly c-conjugate

(64) $$f^{**}: E_1 \to \overline{\mathbb{R}}^1, \quad f^{**}(x) = \sup_{y \in E_2} (c(x,y) - f^*(y)).$$

Then, for any admissible pair (f,g) we have:

(65) $f(x) + g(y) \geq f(x) + f^*(y) \geq f^{**}(x) + f^*(y) \geq c(x,y)$.

Define the <u>equality</u> <u>domains</u> of (f,f^*) by

(66) $E_c f(x) = \{y;\ f(x) + f^*(y) = c(x,y)\}$

$E_c f^*(y) = \{x;\ f(x) + f^*(y) = c(x,y)\}$.

Define the class of <u>c-convex</u> functions

(67) $\Gamma^c(E_1) = \{h: E_1 \to \overline{\mathbb{R}};\ h(x) = \sup_{i\in I} [c(x,y_i) + a_i]$ for some $a_i \in \overline{\mathbb{R}}$, $y_i \in E_2,\ i \in I\}$

$\Gamma^c(E_2) = \{h: E_2 \to \overline{\mathbb{R}};\ h(y) = \sup_{i\in I} [c(x_i,y) + b_i],\ b_i \in \overline{\mathbb{R}},\ x_i \in E_1$, I any index set$\}$.

Elster and Nehse (1974) proved that

a) $f^* \in \Gamma^c(E_2)$, $f^{**} \in \Gamma^c(E_1)$,

b) f^{**} is the largest c-convex function which is majorized by f,

c) $f = f^{**} \Leftrightarrow f \in \Gamma^c(E_1)$.

If $c(x,y) = \langle x,y\rangle$, $x \in Y = E_1$ a locally convex topological vector space, $y \in Y^* = E_2$, then $\Gamma^c(E_1)$ is identical to the class of convex, closed (= lower semicontinuous) functions on Y. From (64) it is clear that in the duality theorem (61) we can restrict to c-convex functions. It is however known that for certain classes of functions the class of c-convex functios is very large, so that in these cases the reduction is not very interesting (cf. [19], [1]).

<u>Theorem</u> 18. For $c \in \mathfrak{L}_m$ with $\int c(x,a)dP_i(x) < \infty$, $i = 1,2$, we have: $P \in M_{\mathfrak{C}}$ is a maximal measure induced by random variables $X \sim P_1$, $Y \sim P_2$, if and only if

(68) $Y \in E_c f(X)$ a.s. for some c-convex $f \in \mathfrak{L}^1(P_1)$ or, equivalently, if and only if $X \in E_c f^*(Y)$.

<u>Proof.</u> If $Y \in E_c f(X)$ a.s. for some c-convex $f \in \mathfrak{L}^1(P_1)$, then for any random variables $\tilde{X} \sim P_1$, $\tilde{Y} \sim P_2$ we have: $Ec(\tilde{X},\tilde{Y}) \leq Ef(\tilde{X}) + Ef^*(\tilde{Y}) = E(f(X) + f^*(Y)) = Ec(X,Y))$, i.e. (X,Y) is an optimal coupling.

There exists a solution (f,g) of the dual problem, $f \in \mathfrak{L}^1(P_1)$, $g \in \mathfrak{L}^1(P_2)$. By (65) we can w.l.g. assume that f is c-convex and $g = f^*$. The converse direction is implied by (62). □

From (68) it is of interest to characterize the equality sets of c-convex functions. For $f: E_1 \to \overline{\mathbb{R}}$ $\varepsilon > 0$, define the <u>ε-c-subdifferential</u>

(69) $\partial_{c,\varepsilon} f(x) = \{y: f(x') - f(x) \geq c(x',y) - c(x,y) - \varepsilon,\ \forall x' \in E_1\}$,

$\partial_c f(x) = \partial_{c,0} f(x)$ the c-subdifferential. The elements of $\partial f(x)$ are called c-subgradients of f in x. There is the following characterization (cf. [19], [27], [1])

(70) $$y \in \partial_c f(x) \Leftrightarrow y \in E_c f(x) \quad \text{(i.e. } f(x) + f^*(y) = c(x,y))$$
$$\Leftrightarrow f(x) - c(x,y) = \inf_{x'} (f(x') - c(x',y))$$
$$(\Leftrightarrow x \in \partial_c f\ (y), \text{ if } f \text{ is c-convex}).$$

If $\partial_{c,\varepsilon} f(x) \neq \emptyset$ for all $0 < \varepsilon \le \varepsilon_0$, then $f(x) = f^{**}(x)$.

Lemma 19. Let $\Phi: E_1 \to E_2$, $\Phi(x) \in \partial_c f(x)$ for $x \in A$, then

(71) $$c(y,\Phi(x)) + c(x,\Phi(y)) \le c(x,\Phi(x)) + c(y,\Phi(y)), \ \forall x,y \in A.$$

Proof. Since $f(y) - f(x) \ge c(y,\Phi(x)) - c(x,\Phi(x))$ and $f(x) - f(y) \ge c(x,\Phi(y)) - c(y,\Phi(y))$, (71) follows by adding these inequalities. □

Remark.

a) If $c(x,y) = -|x-y|^2$, $x,y \in \mathbb{R}^k$, then (71) is equivalent to the monotony of Φ,

(72) $$\langle y - x, \Phi(y) - \Phi(x)\rangle \ge 0.$$

If $\Phi = \nabla f$, f continuous, differentiable, then from (35), this is necessary and sufficient for $\Phi(x) \in \partial_c f(x) = \partial f(x)$.

If f,g,c are differentiable and $\Phi: \mathbb{R}^k \to \mathbb{R}^k$, then the condition that $\Phi(x) \in \partial_c f(x)$ implies that $h(y) := f(y) - f(x) - c(y,\Phi(x)) - c(x,\Phi(x)) \ge 0$ has a minimum in $y = x$ and, therefore,

(73) $$\nabla f(x) = \partial_1 c(x,\Phi(x)).$$

If the differential form $\omega = \partial_1 c(x,\Phi(x)) \cdot dx$ is closed, we obtain

(74) $$f(x) = c_1 + \int_{0\to x} \partial_1 c(x,\Phi(x)) \cdot dx.$$

Similarly,

(75) $$\nabla f^*(\Phi(x)) = \partial_2 c(x,\Phi(x))$$

and if Φ is invertible and $\partial_2 c(x,\Phi(x)) \cdot dx$ is closed, then

(76) $$f^*(y) = c_2 + \int_{\Phi(0)\to y} \partial_2 c(\Phi^{-1}(u),u) \cdot du.$$

With the substitution $v = \Phi^{-1}(u)$, i.e. $du = \Phi'(v)dv$. We define $c_1 + c_2 = c(0,\Phi(0))$; then we obtain

(77) $$f(x) + f^*(\Phi(x)) = c(0,\Phi(0)) + \int_{0\to x} [\partial_1 c(u,\Phi(u) \cdot u + \partial_2 c(u,\Phi(u))\Phi'(u)] \cdot du$$
$$= c(0,\Phi(0)) + \int_{0\to x} d(c(u,\Phi(u))) = c(x,\Phi(x)).$$

Therefore, the condition that $\Phi(x)$ is the c-subgradient of a differentiab function f, is equivalent to

(78) $$f(x) + f^*(y) = f(x) - f(\Phi^{-1}(y)) + f(\Phi^{-1}(y)) + f^*(y) = c(\Phi^{-1}(y),y) + \int_{\Phi^{-1}(y)\to x} \partial_1 c(u,\Phi(u))\cdot du \geq c(x,y),$$

equivalently, to the differential characterization

(79) $$\int_{\Phi^{-1}(y)\to x} [\partial_1 c(u,y) - \partial_1 c(u,\Phi(u))]\cdot du \leq 0, \quad \forall\, x,y.$$

(The case $c(x,y) = -|x-y|^\alpha$, $\alpha > 1$, has been considered in [96].)

As consequence of this discussion we obtain

Theorem 20. If Φ is continuously differentiable, injective and if $\partial_i c(x,\Phi(x))\cdot dx$, $i = 1,2$ is closed, then: $\Phi(x) \in \partial_c f(x)$, $\forall\, x$, for a continuous differentiable function f if and only if (79) holds for all x,y. □

3.2. Generalizations of Young's Inequality

In this section we consider some generalizations of the Young-inequality. Let Φ: $[0,\infty) \to [0,\infty)$ be a Young-function i.e. Φ is right continuous, nondecreasing, $\Phi(0) = 0$ and $\Phi(x) \xrightarrow[x\to\infty]{} \infty$ and define the generalized inverse $\Phi^-(y) = \sup\{x: \Phi(x) \leq y\}$. The Young-inequality states the inequality:

(80) $$xy \leq \int_0^x \Phi(t)dt + \int_0^y \Phi^-(s)ds$$

for all x,y > 0 with equality if and only if $\Phi(x-) \leq y \leq \Phi(x)$ (cf. [6], [57], [17]).

Define for a measure generating function F on $[0,\infty)^2$ and coresponding measure m

(81) $$h_1(x) := m\{(s,t);\ 0 \leq s \leq x,\ 0 \leq t \leq \Phi(s)\} = \int_0^x \Big(\int_0^{\Phi(s)} dF(t|s)\Big)dF_1(s)$$
$$h_2(y) := \int_0^y \Big(\int_0^{\Phi^-(t)} dF(s|t)\Big)dF_2(t),$$

$F(\cdot|\ )$, $F_i(\cdot)$ denote the conditional resp. marginal "distribution" functions.

Theorem 21. For x,y > 0 we have:

(82) $$F(x,y) + F(0,0) \leq (F(x,0) + h_1(x)) + (F(0,y) + h_2(y)).$$

Proof. Define $A = [0,x]\times[0,y]$, $B = \{(s,t);\ 0 \leq s \leq x,\ 0 \leq t \leq \Phi(s)\}$, $C = \{(s,t);\ 0 \leq t \leq y,\ 0 \leq s < \Phi^-(y)\}$. Then

(83) $$A \subset B \cup C \quad \text{and} \quad B \cap C = \emptyset.$$

Therefore, $m(A) = F(x,y) - F(x,0) - F(0,y) + F(0,0) \leq m(B) + m(C) = h_1(x) + h_2(y)$.

□

Remark.

a) The idea of the proof of Theorem 21 is due to Pales (1987) who noted that in the classical geometric proof one can use more general measures.

If $m = f\lambda_+^2$, then we obtain more explicitely:

(84) $$h_1(x) = \int_0^x \left(\int_0^{\Phi(s)} f(s,t)dt \right) ds = \int_0^x (\partial_1 F(s,\Phi(s)) - \partial_1 F(s,0))ds$$
$$= \int_0^x \partial_1 F(s,\Phi(s)) - F(x,0) + F(0,0)$$

and

$$h_2(y) = \int_0^y \partial_2 F(\Phi^-(t),t)dt - F(0,y) + F(0,0),$$

where the partial derivatives exist a.s. w.r.t. the Lebesgue measure. Therefore, from (82)

(85) $$F(x,y) \le F(0,0) + \int_0^x \partial_1 F(s,\Phi(s))ds + \int_0^y \partial_2 F(\Phi^-(t),t)dt.$$

This inequality is due to Pales (1987) for $F \in C^2$ with $\partial_1\partial_2 F(s,t) \ge 0$.

Example. Let for $\alpha > 1$, $F(x,y) = -|x-y|^\alpha$, $x,y \in \mathbb{R}_+^1$, then $\partial_1\partial_2 F(x,y) = \alpha(\alpha-1)|x-y|^{\alpha-2} \ge 0$. Therefore, by (85) we obtain the inequality

(86) $$|x-y|^\alpha \ge \alpha \int_0^x |s-\Phi(s)|^{\alpha-1} sg(s-\Phi(s))ds$$
$$+ \alpha \int_0^y |t-\Phi^-(t)|^{\alpha-1} sg(t-\Phi^-(t))dt,$$

where $sg(x) = \begin{cases} 1 & \\ 0 & \text{if } x \overset{>}{\underset{<}{=}} 0 \\ -1 & \end{cases}$ and Φ is a Young function. An analytical derivation of (86) has been given in [96]. A consequence of (86) and (62) is the wellknown fact that random variables (X,Y) with $\Phi(X-) \le Y \le \Phi(X)$ and Φ a Young function are optimal couplings w.r.t. the distance $c(x,y) = |x-y|^\alpha$. □

We next derive an extension of (82) to the case of the whole real line.

Theorem 22. Let $\Phi: \mathbb{R}^1 \to \mathbb{R}^1$ be nondecreasing, right continuous. Let F be the generating function of a finite measure, $F(x,y) = P(X \le x, Y \le y)$, then

(87) $$F(x,y) \le f(x) + g(y), \text{ where } f(x) = \int_{-\infty}^x P(Y \le \Phi(s)|X=s)dP^X(s) \quad \text{and}$$
$$g(y) = \int_{-\infty}^x P(X < \Phi^-(s)|Y=s)dP\;\;(s).$$

Proof. $F(x,y) = P(X \le x, Y \le \Phi(X) \wedge y) + P(X \le x, Y \le y, Y > \Phi(X)) \le$
$P(X \le x, Y \le \Phi(X)) + P(X < \Phi^{-}(Y), Y \le y) = \int_0^x P(Y \le \Phi(X)|X = s)dP^{X}(s) +$
$\int_0^x P(X > \Phi^{-}(Y)|Y = s)dP^{Y}(s) = f(x) + g(y).$ □

In (87) we have equality, iff

(88) $$\Phi(X-) \le Y \le \Phi(X) \text{ a.s.}$$

(85), (87) imply optimal coupling results and Fréchet bounds for Δ-monotone (resp. L-superadditive) functions (cf. (21)).

An extension of the Young inequality to n-variables is the Oppenheim inequality. Let $f_i: [0,\infty) \to [0,\infty)$ be Young-functions, $1 \le i \le n$, then:

(89) $$\prod_{i=1}^n f_i(t_i) \le \sum_{i=1}^n \int_0^{t_i} (\prod_{j \ne i} f_j) df_i .$$

This inequality was used in Gaffke and Rüschendorf (1981) to determine $M_{\mathfrak{G}}(\varphi)$ for $\varphi(x) = \prod_{i=1}^n x_i$ and simple marginals $P_1, \ldots, P_n$. For the literature cf. Oppenheim (1927), Cooper (1927), and Dankert and König (1967).

Consider the curve $y(t) = (f_1(t), \ldots, f_n(t))$, $t \ge 0$ and the points $P_i = (f_j(t_i))_{1 \le j \le n}$, $1 \le i \le n$, $A := (f_i(t_i))_{1 \le i \le n}$. Define, furthermore,

(90) $$V_i := \{x \in \mathbb{R}_+;\ x_i \le f_i(t_i),\ x_j \le f_j(f_i^{-}(x_i)),\ j \ne i\} \quad \text{and}$$
$$V := [0,A] = [0,f_1(t_1)] \times \ldots \times [0,f_n(t_n)] .$$

Theorem 23. (Generalized Oppenheim Inequality)
Let m be a Radon measure on $\mathbb{R}_+^n$ and define $\varphi(t_1,\ldots,t_n) := m([0,A])$, $h_i(t_i) = m(V_i)$, $1 \le i \le n$, then:

(91) $$\varphi(t_1,\ldots,t_n) \le \sum_{i=1}^n h_i(t_i) .$$

Proof. The proof follows from the inclusion $V = [0,A] \subset \bigcup_{i=1}^n V_i$ (cf. [65]). □

With $m = \lambda^n$, the Lebesgue measure, we have from (91)
$\prod_{i=1}^n f_i(t_i) \le \sum_{i=1}^n \int_0^{f_i(t_i)} \prod_{j \ne i} f_j(f_i^{-}(s))ds = \sum_{i=1}^n \int_0^{t_i} (\prod_{j \ne i} f_j(u))df_i(u)$, the Oppenheim inequality. □

For finite measures m we can extend (91) to $\mathbb{R}^n$.

Theorem 24. Let $m = P^{(U_1,\ldots,U_n)}$ be a finite measure on $\mathbb{R}^n$ with generating function F and define: $\varphi(t_1,\ldots,t_n) := F(f_1(t_1),\ldots,f_n(t_n))$,

$$h_i(t_i) := P(U_i \le f_i(t_i),\ U_j \le f_j \circ f_i^-(U_i),\ j \ne i) = \int_{-\infty}^{t_i} P(U_j \le f_j(u),\ j \ne i \,|\, U_i = f_i(u))\, dP^{f_i^-(U_i)}(u),$$

then

$$\varphi(t_1,\dots,t_n) \le \sum_{i=1}^{n} h_i(t_i). \tag{92}$$

<u>Proof.</u> $\varphi(t_1,\dots,t_n) = P(U_i \le f_i(t_i),\ 1 \le i \le n) = P(f_i^-(U_i) \le t_i,\ 1 \le i \le n) \le$ $\sum_{i=1}^{n} P(f_j^-(U_j) \le t_j \wedge f_i^-(U_i),\ 1 \le j \le n) \le \sum_{i=1}^{n} P(f_i^-(U_i) \le t_i,\ f_j^-(U_j) \le f_i^-(U_i),\ \forall j \ne i) \le$ $\sum_{i=1}^{n} P(U_i \le f_i(t_i),\ U_j \le f_j \circ f_i^-(U_i),\ \forall j \ne i) = \sum_{i=1}^{n} \int_{-\infty}^{f_i(t_i)} P(U_j \le f_j \circ f_i^-(s), \forall j \ne i \mid U_i = s)\, dP^{U_i}(s) =$ $\sum_{i=1}^{n} \int_{-\infty}^{t_i} P(U_j < f_j(u),\ j \ne i \mid U_i = f_i(u))\, dP^{f_i^-(U)}(u)$. □

4. <u>Some Statistical Applications and Problems</u>

4.1. <u>Marginal Sufficiency</u>

Let $\mathfrak{P}$ be a dominated set of product measures on $(E,\mathfrak{A}) = \prod_{i=1}^{n}(E_i,\mathfrak{A}_i)$ and define $T: (E,\mathfrak{A}) \to (Y,\mathfrak{B})$ to be <u>marginally sufficient</u> for $\mathfrak{P}$, if for all $1 \le i \le n$ and $\varphi \in B(E,\mathfrak{A})$, $\varphi = \varphi(x_i)$, there exists $\tilde{\varphi} \in B(E,\sigma(T))$ with

$$\tilde{\varphi} = E_P(\varphi|T)\ [P],\ \forall P \in \mathfrak{P}. \tag{93}$$

Huzurbazar proposed the following conjecture.

<u>Huzurbazar conjecture:</u>

(94) Marginal sufficiency of T implies sufficiency

(i.e. (93) is true for any $\varphi = \varphi(x_1,\dots,x_n) \in B(E,\mathfrak{A})$).

The first published proof of (94) was given by Sudakov (1979) in the case of equivalent measures (cf. also [53]). The idea of Sudakov's proof is related to some marginal problems. The idea is the following (cf. [105], p. 154 - 160). Let for $P = \otimes P_i$, $Q = \otimes Q_i \in \mathfrak{P}$, $f = \Pi f_i$, $g = \Pi g_i$ be densities w.r.t. a dominating measure $\mu = \otimes \mu_i$. Let $h_i = g_i / f_i$, $\varphi(x) := (\ln h_1(x_1),\dots,\ln h_n(x_n))$ and let P_y, Q_y denote the conditional distributions of P, Q given $T = y$. If T is marginally sufficient, then $\tilde{P}_y := P^{\varphi|T=y} = (P_y)^\varphi$ and $\tilde{Q}_y := Q^{\varphi|T=y} = (Q_y)^\varphi$, $y \in Y$, are probability measures on $\mathbb{R}^n$ with identical marginals. Using $\frac{dQ_y}{dP_y}(x) = \frac{dP^T}{dQ^T}(y)\, \Pi h_i(x_i) = \frac{dP^T}{dQ^T}(y) \exp(\Sigma \ln h_i(x_i))$ one concludes:

$$\frac{d\tilde{Q}_y}{d\tilde{P}_y}(z) = \frac{dP^T}{dQ^T}(y) \exp(\Sigma z_i). \tag{95}$$

Let U, V be orthogonal, nonnegative measures with $P_y - Q_y = U - V$, then

1. U,V have identical marginals,
2. $U(\Sigma x_k \le \ell) = 0$, $V(\Sigma x_k \ge \ell) = 0$ for some ℓ (namely $\ell = -\ell n \frac{dP^T}{dQ^T}(y)$),
3. $\int |x_k| dU < \infty$, $1 \le k \le n$.

To establish 3. is the most involved part of Sudakov's proof. It is easy to see that

(96) 1., 2., 3. implies that $U = V = 0$.

As consequence: $\tilde{P}_y = \tilde{Q}_y$ and, therefore, a standard argument from sufficiency theory implies that T is sufficient for {P,Q}. Since this holds for any pair P, Q, T is sufficient for $\mathfrak{P}$.

The following interesting example of Sudakov shows that the difficult moment condition 3. cannot be omitted.

Example. Let $\varphi: \mathbb{Z}^3 \to \mathbb{Z}^3$, $\varphi(x) = -x$ and define probability measures P,Q by

(97) $$\frac{3}{4} P = \frac{1}{8} S\varepsilon_{(1,1,-1)} + \sum_{k=2}^{\infty} \frac{1}{2^{k+2}} S\varepsilon_{(2^k-1,1-2^{k-1},1-2^{k-1})}$$
$$Q := P^{\varphi},$$

where $S\varepsilon_{(1,1,-1)} = \varepsilon_{(1,1,-1)} + \varepsilon_{(1,-1,1)} + \varepsilon_{(-1,1,1)}$, ε_x the one point measure in x. Then the marginals of P, Q are identical and equal to $\frac{1}{3}(\varepsilon_{-1} + \varepsilon_1) + \frac{4}{3} \sum_{k=2}^{\infty} 2^{-(k+2)}(\varepsilon_{2^k-1} + \varepsilon_{(1-2^k)})$ and $P\{x \in \mathbb{Z}^3;\ \Sigma x_i = 1\} = 1$, while $Q\{\Sigma x_i = -1\} = 1$. □

Let more generally $\mathfrak{A}_1 \subset \dots \subset \mathfrak{A}_n \subset \mathfrak{A}$ be an increasing sequence of σ-algebras and $P, Q \in M^1(E,\mathfrak{A})$. Let $Q \ll P$, $P_k = P/\mathfrak{A}_k$, $Q_k = Q/\mathfrak{A}_k$, $L_k = \frac{dQ_k}{dP_k}$ and $f_k := L_k / L_{k-1}$.

Theorem 25. (Generalized Huzurbazar conjecture, cf. [95])
If $T: (E,\mathfrak{A}) \to (Y,\mathfrak{B})$ is partially sufficient for $\sigma(f_k)$, $1 \le k \le n$, then T is sufficient for $\{P_n, Q_n\}$.

The proof uses the following two lemmas:

Lemma 1. (cf. [78], Prop. 6)
If $P_i \in M^1(\mathbb{R}^1)$, $1 \le i \le n$, $P,Q \in M(P_1,\dots,P_n)$, then:

(98) $P \le_{st} Q$ implies $P = Q$. □

Lemma 2. (cf. Simons (1980))

For any sub-σ-algebra $\mathfrak{B} \subset \mathfrak{A}$ and $P,Q \in M^1(E,\mathfrak{A})$ the conditional distributions of $L = (L_1,\ldots,L_n)$ w.r.t. $\mathfrak{B}$ satisfy:

$$(99) \qquad P^{L|\mathfrak{B}} \le_{st} Q^{L|\mathfrak{B}} . \qquad \square$$

If T is partially sufficient for $\sigma(f_k)$, $1 \le k \le n$, then $P_{\mathfrak{B}}^{(f_1,\ldots,f_n)}, Q_{\mathfrak{B}}^{(f_1,\ldots,f_n)}$ have the same marginals, where $P_{\mathfrak{B}}$, $Q_{\mathfrak{B}}$ are the conditional distributions. Then by a generalization of the a.s. representation theorem of stochastic orders to Markov kernels, one obtains versions $X_{\mathfrak{B}}$, $Y_{\mathfrak{B}}$ of the distribution in (99) such that $X_{\mathfrak{B}} \le Y_{\mathfrak{B}}$ a.s. From these versions one can construct versions $\tilde{X}_{\mathfrak{B}}$, $\tilde{Y}_{\mathfrak{B}}$ of the distributions of $P_{\mathfrak{B}}^{(f_1,\ldots,f_n)}, Q_{\mathfrak{B}}^{(f_1,\ldots,f_n)}$ such that $\tilde{X}_{\mathfrak{B}} \le \tilde{Y}_{\mathfrak{B}}$ a.s. Therefore, from (98) one obtains $P_{\mathfrak{B}}^{(f_1,\ldots,f_n)} = Q_{\mathfrak{B}}^{(f_1,\ldots,f_n)}$. This implies that $\mathfrak{B}$ is sufficient, $\mathfrak{B} = \sigma(T)$.

4.2. Optimal Combination of Tests of Marginals

Let P_i, $Q_i \in M^1(E_i,\mathfrak{A}_i)$, $1 \le i \le n$, and consider the testproblem with hypothesis $\Theta_o = M(P_1,\ldots,P_n)$ and alternative $\Theta_1 = M(Q_1,\ldots,Q_n)$. In a practical problem this means e.g. that one measures n components and has for each components the simple alternatives $\{P_i\}$, $\{Q_i\}$ but does not know anything about the dependence structure of the measurements. The question then is the following: Is it possible to achieve a better test Θ_o, Θ_1 then to take the test for that component which allows for a certain test level α the highest power? What is the optimal combination of the marginal tests?

The answer to this problem was given in [85] w.r.t. the maximin criterio
We consider the tests of level α

$$(100) \qquad \Phi_\alpha(\Theta_o) = \{\varphi \in \Phi : E_P\varphi \le \alpha, \ \forall P \in M(P_1,\ldots,P_n) = \Theta_o\}$$

and the maximinrisk

$$(101) \qquad \beta(\alpha,\Theta_o,\Theta_1) = \sup_{\varphi \in \Phi_\alpha(\Theta_o)} \inf_{P \in \Theta_1} E_P\varphi .$$

Let for two finite measures P,Q on $(E,\mathfrak{A})$

$$(102) \qquad d_v(P,Q) = \sup\{P(A) - Q(A);\ A \in \mathfrak{A}\}$$
$$d_v(\mathfrak{P},\mathfrak{Q}) = \inf\{d_v(P,Q);\ P \in \mathfrak{P},\ Q \in \mathfrak{Q}\}$$

for subsets $\mathfrak{P},\mathfrak{Q} \subset M(E,\mathfrak{A})$ and define

$$(103) \qquad h_\alpha(x) := \alpha x + \max_{1 \le i \le n} d_v(Q_i, xP_i), \ x \ge 0 .$$

Theorem 26. (cf. [85])

Let $\alpha \in (0,1)$ and let x^* be a minimum point of h_α, then:

a) $\beta(\alpha,\Theta_o,\Theta_1) = h_\alpha(x^*)$.

b) If $P \in \Theta_o$, $Q \in \Theta_1$ satisfy

(104) $$d_v(Q, x^* P) = d_v(\Theta_1, x^* \Theta_o),$$

then there exists a LQ-test φ^* for $(\{P\},\{Q\})$ with critical value x^* such that φ^* is a maximin level α-test, i.e.

(105) $$\varphi^* \in \Phi_\alpha(\Theta_o) \quad \text{and} \quad \inf_{P \in \Theta_1} E_P \varphi^* = \beta(\alpha,\Theta_o,\Theta_1). \qquad \square$$

The proof uses the following lemma.

Lemma 3. $\forall x \geq 0$ holds:

(106) $$d_v(M(Q_1,\dots,Q_n), x\, M(P_1,\dots,P_n)) = \max_{1 \leq i \leq n} d_v(Q_i, x\, P_i). \qquad \square$$

Minimal pairs can explicitely be determined. Furthermore, the proof uses a characterization of maximintests given by Baumann (1968).

One can not improve the best test of single marginals if e.g.

(107) $$d_v(Q_1, x\, P_1) = \max_j d_v(Q_j, x\, P_j), \ \forall x \geq 0.$$

But in other cases one obtains a considerable improvement. Some related results with additional restrictions on the hypotheses have been discussed in [93].

An alternative interpretation of Theorem 26 is in terms of robustness. If $M(P_1,\dots,P_n)$ is considered as a neighbourhood of $P_1 \otimes \dots \otimes P_n$, then for a test φ, $M_{\mathfrak{C}}(\varphi)$ is its robust level and Theorem 26 constructs an optimal robust test.

4.3. Optimal Estimators in Marginal Models

We consider the construction of minimum variance unbiased estimators (MVU) in the model $\mathfrak{P} = M(P_1,\dots,P_k)$ for certain functions $g : \mathfrak{P} \to \mathbb{R}^1$. The general question is the following: How can one use the knowledge of the marginals in order to construct better estimators than those in the model without this knowledge?

Let D_o be the set of all unbiased estimators of zero, let $\overline{P} := \bigotimes_{i=1}^k P_i$ and let D_g denote the unbiased estimators of g.

Theorem 27. (cf. [89])

a) $D_o = F_1 := \{\sum_{i=1}^{k} f_i(x_i);\ f_i \in \mathfrak{L}^1(P_i),\ \int f_i dP_i = 0,\ 1 \le i \le k\}$.

b) If $P \in \mathfrak{P}$ and $d \in D_g \cap L^2(P)$, then

(108) $$d^* := d - E_P(d|\overline{F}_2^P)$$

is MVU for g in P, where $\overline{F}_2^P$ denotes the closure of $F_2 = \{\sum f_i(x_i);\ f_i \in \mathfrak{L}^2(P_i),\ \int f_i dP_i = 0\}$ in $\mathfrak{L}^2(P)$.

c) If $d \in D_g$, then

(109) $$d^* := d - \sum_{i=1}^{k} \int d\ d \bigotimes_{j \neq i} P_j + k \int d\ d\overline{P}$$

is MVU for g in $\overline{P}$.

The projections occuring in Theorem 27 can be calculated in some cases while in the general case an approximative solution based on the alternating projection theorem is known (cf. [84]).

In the case of n independent observations the underlying model is $\mathfrak{P}^n = \{P^n;\ P \in \mathfrak{P}\}$ and the corresponding optimal estimator is given by

(110) $$d_n^*(x_1,\dots,x_n) = \frac{1}{n} \sum_{i=1} d^*(x_i).$$

An estimator sequence for a differentiable functional g, which is asymptotically optimal on the whole model or a subset $\mathfrak{P}_o \in \mathfrak{P}$, should have the stochastic expansion

(111) $$\sqrt{n}\ (d_n^*(x_1,\dots,x_n) - g(P)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_P(x_i) + o_{P^n}(1),\ P \in \mathfrak{P}_o,$$

where g_P lies in the tangent cone $T(P,\mathfrak{P})$, the set of all derivatives (tangent vectors) of L^2-differentiable path's in $\mathfrak{P}$ through P. $T(P,\mathfrak{P})$ can be shown to be identical to

(112) $$T(P,\mathfrak{P}) = (F_2)^{\perp P},$$

the orthogonal complement of F_2 in $L^2(P)$ (cf. [90], [92]). The stochastic expansion in (111) implies that g_P is a gradient of g and since $g_P \in T(P,\mathfrak{P})$, it is the canonical gradient.

In a recent paper Bickel, Ritov and Wellner (1988) succeeded to construct an estimator sequence with this property on the subset $\mathfrak{P}_\alpha \subset \mathfrak{P}$, $\alpha > 0$, $k = 2$, consisting of 'positive dependent' measures P with

(113) $$P(A \times B) \ge \alpha P_1(A) P_2(B),\ \ \forall A,B.$$

References

[1] Balder, S. J.: An extension of duality-stability relations to non-convex optimization problems. SIAM J. Contr. Opt. 15 (1977), 320 - 343

[2] Baumann, V.: Eine parameterfreie Theorie der ungünstigsten Verteilungen. Z. Wahrscheinlichkeitstheorie verw. Geb. 11 (1968), 40 - 60

[3] Bertino, S.: Su di una sottoclasse della classe di Fréchet. Statistica 28 (1968), 511 - 542

[4] Bertino, S.: Sulla distanza tra distribuzioni. Pubbl. Ist. Calc. Prob. Univ. Roma, n. 82 (1968)

[5] Bickel, P. J., Ritov, Y., and Wellner, J. A.: Efficient estimation of a probability measure P with known marginal distributions. Preprint, 1988

[6] Boas, R. P., and Marcus, M. B.: Inverse functions and integration by parts. Amer. Math. Monthly 81 (1974), 760 - 761

[7] Brown, J. R.: Approximation theorems for Markov operators. Pacific J. Math. 16 (1966), 13 - 23

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

[9] Cooper, R.: Notes on certain inequalities. J. London Math. Soc. 2 (1927), 17 - 21

[10] Csiszar, I.: I-divergence geometry of probability distributions and minimization problems. Ann. Prob. 3 (1975), 146 - 158

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

[12] Dall'Aglio, G.: Sulla compatibilita delle funzioni di ripartizione doppia. Rendiconti di Math. 18 (1959), 385 - 413

[13] Dall'Aglio, G.: Les fonctions extrêmes de la classe de Frêchet à 3 dimensions. Publ. Inst. Stat. Univ. Paris, IX (1960), 175 - 188

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

[15] Dankert, G., and König, H.: Über die Höldersche Ungleichung in Orlicz-Räumen. Arch. Math. 118 (1967), 61 - 75

[16] Darroch, J. N., Lauritzen, S. L., and Speed, T. P.: Markov fields and log-linear interaction models for contingency tables. Ann. Statist. 8 (1980), 522 - 539

[17] Diaz, J. B., and Metcalf, F. T.: n analytic proof of Young's inequality. Amer. Math. Monthly 77 (1970), 603 - 609

[18] Dobrushin, R. L.: Prescribing a system of random variables by conditional distributions. Theory Prob. Appl. 15 (1970), 458 - 486

[19] Dolecki, S., and Kurcyusz, St.: On Φ-convexity in extremal problems. SIAM J. Control. Optim. 6 (1978), 277 - 300

[20] Douglas, R. G.: On extremal measures and subspace density. Michigan Math. J. 11 (1964), 243 - 246

[21] Dowson, C. D., and Landau, B. U.: The Fréchet distance between multivariate normal distributions. J. Multivar. Anal. 12 (1982), 450 - 455

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

[23] Dudley, R. M.: Probability and Metrics. Aarhus Univ., Aarhus, 1976

[24] Dudley, R. M.: Real Analysis and Probability. Wadsworth and Brooks/ Cole, 1989

[25] Dunford, N., and Schwartz, J. T.: Linear Operators, Part I. Interscience Publishers, 1967

[26] Edwards, D. A.: On the existence of probability measures with given marginals. Ann. Inst. Fourier 28 (1978), 53 - 78

[27] Elster, K. H., and Nehse, R.: Zur Theorie der Polarfunktionale. Math. Operationsf. u. Statist. 5 (1974), 3 - 21

[28] Fernique, X.: Sur le théorème de Kantorovitch-Rubinstein dans les espaces polonais. Lecture Notes in Mathematics 850, pp. 6 - 10. Springer, 1981

[29] Frank, M. J., Nelsen, R. B., and Schweizer, B.: Best possible bounds for the distribution of a sum - a problem of Kolmogorov. Prob. Theory Rel. Fields 74 (1987), 199 - 211

[30] Fréchet, M.: Sur les tableaux de corrélation dont les marges sont donnees. Annales de l'Universite de Lyon, Sciences 4 (1951), 13 - 84

[31] Fréchet, M.: Les Tableaux de corrélation dont les marges et des bornes sont donnees. Annales Univ. Lyon, Sciences 20 (1957), 13 - 31

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

[33] Fréchet, M.: Sur les tableaux de corrélation dont les marges et des bornes sont donnees. Revue Inst. Int. de Statistique 28 (1960), 10 - 32

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

[35] Gaffke, N., and Rüschendorf, L.: On the existence of probability measures with given marginals. Statistics & Decisions 2 (1984), 163 - 174

[36] Gini, C.: Di una misura della dissomiglianza tra due gruppi di quantità e delle sue applicazioni allo studio delle relazioni statistiche. Att. R. Ist. Veneto Sc. Lettere, Art 74 (1914), 185 - 213

[37] Givens, C. R., and Shortt, R. M.: A class of Wasserstein metrics for probability distributions. Manuscript, Dept. Math. Comp. Sci. Michigan Tech. Univ., Houghton, MI, 1984

[38] Hammersley, I. M., and Handscomb, D. C.: Monte Carlo Methods. Meth London, 1964

[39] Hansel, G., and Troallic, J. P.: Mesures marginales et théorème de Ford-Fulkerson. Z. W.-theorie verw. Gebiete 43 (1978), 245 - 251

[40] Hoeffding, W.: Masstabinvariante Korrelationstheorie. Sem. Math. Inst. Univ. Berlin 5 (1950), 181 - 233

[41] Huang, J. S., and Kotz, S.: Correlation structure in iterated Farlie-Gumbel-Morgenstern distributions. Biometrika 71 (1984),

[42] Ivanov, E. H., and Nehse, R.: Relations between generalized concepts of convexity and conjugacy. Math. Operationsforsch. Statist., Ser. Opt. 13 (1982), 9 - 18

[43] Johnson, N. L., and Kotz, S.: On some generalized Farlie-Gumbel-Morgenstern distributions. Comm. Statistics 4 (1975), 415 - 427

[44] Kamae, T., Krengel, U., and O'Brien, G. L.: Stochastic inequalities on partially ordered spaces. Ann. Prob. 5 (1977), 899 - 912

[45] Kantorovic, L. V., and Rubinstein, G. S.: On a space of completely additive functions (in Russian). Vestnik Leningrad Univ. 13 (1958), 52 - 59

[46] Kellerer, H. G.: Verteilungsfunktionen mit gegebenen Marginalverteilungen. Z. Wahrsch. 3 (1964), 247 - 270

[47] Kellerer, H. G.: Duality theorems for marginal problems. z. Wahrsch. 67 (1984), 399 - 432

[48] Kellerer, H. G.: Measure theoretic versions of linear programming. Preprint, 1987

[49] Kimeldorf, G., and Sampson, A.: Monotone dependence. Ann. Statist. 6 (1978), 895 - 903

[50] Klein Haneveld, W. K.: Robustness against PERT: An application of duality and distributions with known marginals. Preprint, 1984

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

[52] Kotz, S., and Johnson, N.: Propriétés de dépendence des distributions iterees generalisees a deux variables Farlie-Gumbel-Morgenstern. C. R. Acad. Sc. Paris 285 (1977), 277 - 280

[53] Kudo, H.: On marginal sufficiency. Statistics & Decisions 4 (1986), 301 - 320

[54] Lai, T. L., and Robbins, M.: Maximally dependent random variables. Proc. Nat. Acad. Sci. USA, 73 (1976), 286 - 288

[55] Lancaster, H. O.: Correlation and complete dependence of random variables. Ann. Math. Statist. 34 (1963), 1315 - 1321

[56] Lauritzen, S. L., and Wermuth, N.: Graphical models for associations between variables. Ann. Statist. 17 (1989), 31 - 57

[57] Lembcke, J.: Gemeinsame Urbilder endlich additiver Inhalte. Math. Ann. 198 (1972), 239 - 258

[58] Levin, V. L., and Malyutin, A. A.: The mass transfer problem with discontinuous cost function and a mass setting for the problem of duality of convex extremum problems. Uspekhi Mat. Nauk 34 (1979), 3 - 68 (in Russian)

[59] Losonczi, L.: Inequalities of Young-type. Monatsh. Math. 97 (1984), 125 - 132

[60] Luschgy, H., and Thomsen, W.: Extreme points in the Hahn-Banach-Kantorovic setting. Pacific J. Math. 105 (1983), 387 - 398

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

[62] Mardia, K. V.: Families of Bivariate Distributions. Hafner, Darim, 1970

[63] Mosler, K. C.: Entscheidungsregeln bei Risiko: Multivariate stochastische Dominanz. Lecture Notes in Economics and Math. Systems 204. Springer, 1982

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

[65] Oppenheim, A.: Note on Cooper's generalization of Young's inequality. J. London Math. Soc. 2 (1927), 21 - 23

[66] Pales, Z.: A Generalization of Young's inequality. In: Inequalities V. Ed. E. Walter (1987)

[67] Plackett, R. L.: A class of bivariate distributions. J. Am. Stat. Assoc. 60 (1965), 516 - 522

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

[69] Rachev, S. T.: The Monge Kantorovich mass transference problem and its stochastic applications. Theroy Prob. Appl. 24 (1985), 647 - 671

[70] Rachev, S. T., and Shortt, R. M.: Duality theorems for Kantorovic-Rubinstein and Wasserstein functions. Preprint, 1989

[71] Rachev, S. T., and Rüschendorf, L.: A transformation property of minimal metrics. To appear in: Theory Prob. Appl., 1989

[72] Rachev, S. T., Rüschendorf, L., and Schief, A.: On the construction of almost surely convergent random variables. Preprint: Angew. Math. und Informatik 10 (1988)

[73] Rachev, S. T., and Rüschendorf, L.: A counterexample to a.s. constructions. Stat. Prob. Letters 9 (1990), 307 - 309

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

[75] Rüschendorf, L.: Vergleich von Zufallsvariablen bzgl. integralinduzierter Halbordnungen. Habilitationsschrift, Aachen, 1979

[76] Rüschendorf, L.: Inequalities for the expectation of Δ-monotone functions. Z. Wahrsch. 54 (1980), 341 - 354

[77] Rüschendorf, L.: Ordering of distributions and rearrangement of functions. Ann. Probab. 9 (1980), 276 - 283

[78] Rüschendorf, L.: Stochastically ordered distributions and monotonicity of the OC of an SPRT. Math. Operationsf., Statistics 12 (1981), 327 - 338

[79] Rüschendorf, L.: Sharpness of Frechet bounds. Z. Wahrsch. 57 (1981), 293 - 302

[80] Rüschendorf, L.: Random variables with maximum sums. Adv. Appl. Prob. 14 (1982), 623 - 632

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

[82] Rüschendorf, L.: On the multidimensional assignment problem. Methods of OR 47 (1983), 107 - 113

[83] Rüschendorf, L.: On the minimum discrimination information theorem. Statistics and Decisions, Suppl. Issue No. 1 (1984), 263 - 283

[84] Rüschendorf, L.: Projections and iterative procedures. In: Proceedings Sixth Intern. Symp. Mult. Anal., Pittsburgh. Ed. P. R. Krishnaiah (1985), 485 - 593

[85] Rüschendorf, L.: Robust tests against dependence. Prob. Math. Statistics 6 (1985), 1 - 10

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

[87] Rüschendorf, L.: Construction of multivariate distributions with given marginals. Ann. Inst. Stat. Math. 37 (1985), 225 - 233

[88] Rüschendorf, L.: Monotonicity and unbiasedness of tests via a.s. constructions. Statistics 17 (1986), 221 - 230

[89] Rüschendorf, L.: Unbiased estimation in nonparametric classes of distributions. Statistics and Decisions 5 (1987), 89 - 104

[90] Rüschendorf, L.: Unbiased estimation and local structure. Proceedings 5th Pannonian Symposium in Visegrad (1985), 295 - 306

[91] Rüschendorf, L.: Projections of probability measures. Statistics 18 (1987), 123 - 129

[92] Rüschendorf, L.: Unbiased estimation of von Mises functionals. Statist. Prob. Letters 5 (1987), 287 - 292

[93] Rüschendorf, L.: Maximintests for neighbourhoods caused by dependence. In: Proceedings 1st World Congress of the Bernoulli Soc., Tashkent, 1986

[94] Rüschendorf, L., and Rachev, S. T.: A characterization of random variables with minimum L^2-distance. J. Mult. Analysis 32 (1990), 4[illegible] - 54

[95] Rüschendorf, L.: Conditional stochastic order and partial sufficiency. To appear in: Adv. Appl. Prob., 1989

[96] Rüschendorf, L.: Bounds for distributions with multivariate marginals. To appear in: Proceedings: Stochastic Order and Decisions under Risk. Ed. K. Mosler and M. Scarsini, 1989

[97] Salvemini, T.: Nuovi procedimenti per il calcolo degli indici di dissomiglianza e di connessione. Statistica (1949), 3 - 26

[98] Scarsini, M.: Lower bounds for the distribution function of a k-dimensional n-extendible exchangeable process. Statist. Prob. Letters 3 (1985), 57 - 62

[99] Scarsini, M.: Copulae of probability measures on product spaces. To appear in: J. Mult. Anal., 1989

[100] Shortt, R. M.: Combinatorial methods in the study of marginal problems over separable spaces. J. Math. Anal. 97 (1983), 462 - 479

[101] Simons, G.: Extensions of the stochastic ordering property of likelihood ratios. Ann. Statist. 8 (1980), 833 - 839

[102] Snijders, T. A. B.: Antithetic variates for Monte Carlo estimation of probabilities. Statistics Neerlandica 38 (1984), 1 - 19

[103] Stoyan, D.: Comparison Methods for Queues and other Stochastic Models. Wiley, 1983

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

[105] Sudakov, V. N.: Geometric problems in the theory of infinite dimensional probability distributions. Proc. Steklov Institute 141 (1979), 1 - 178

[106] Szulga, A.: On the Wasserstein metric. In: Transactions of the 8th Prague Conf. on Inform. Theory, Statist. Decision Funct., and Random Processes, Prague, 1978, v. B. Akademia, Praha, pp. 267 - 273

[107] Szulga, A.: On minimal metrics in the space of random variables. Theory Prob. Appl. 27 (1982), 424 - 430

[108] Tchen, A. H.: Inequalities for distributions with given marginals. Ann. Probab. 8 (1980), 814 - 827

[109] Vallender, S. S.: Calculation of the Wasserstein distance between probability distributions on the line. Theory Prob. Appl. 18 (1973), 784 - 786

[110] Vorobev, N. N.: Consistent families of measures and their extensions. Theory Prob. Appl. 7 (1962), 147 - 163

[111] Warmuth, W.: Marginal Fréchet-bounds for multidimensional distribution functions. Statistics 19 (1988), 283 - 294

[112] Whitt, W.: Bivariate distributions with given marginals. Ann. Statist. 4 (1976), 1280 - 1289