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 \( E = E_1 \times \ldots \times E_n \), \( \mathcal{X} = \mathcal{X}_1 \circ \ldots \circ \mathcal{X}_n \) be a finite product of measure spaces. Let \( \mathcal{G} \subset 2^{\{1, \ldots, n\}} \), the system of all subsets of \( \{1, \ldots, n\} \) with \( \bigcup_{J \in \mathcal{G}} J = \{1, \ldots, n\} \) and let for \( J \in \mathcal{G} \), \( P_J \in M (\prod_{j \in J} E_j) \) be a consistent system of probability measures on \( \pi_J (E) = \prod_{j \in J} E_j \triangleq E_J \), where \( \pi_J \) denotes the J-projection from \( E \) to \( E_J \). Define the marginal model \( M_{\mathcal{G}} \):
\( M_\mathcal{E} = M(P_J, J \in \mathcal{E}) = \{ P \in M(E, \mathcal{X}); P^{\prod J} = P_J, J \in \mathcal{E} \} \)

to be the set of all probability measures on \( E \) with marginals \( P_J = P^{\prod J} \) of
the \( J \)-components, \( J \in \mathcal{E} \).

There are some different type of problems of interest in marginal models and related to Frechet-bounds. The marginal problem is the questi-
on, whether \( M_\mathcal{E} \neq \emptyset \). It was shown by Vorobev (1962), Kellerer (1964), that the property "consistency of \( (P_J)_{J \in \mathcal{E}} \) implies \( M_\mathcal{E} \neq \emptyset " \) is a purely combinator-
torial (graphtheoretic) property and is equivalent to the nonexistence of "cycles" in \( \mathcal{E} \). Systems \( \mathcal{E} \) with this property are called decomposable resp.
simplicial complex (in [100]). Some related existence problems are investiga-
gated in [46], [26], [39], [35]. For non-regular systems the marginal pro-
blem is generally not easy to decide except in cases, where explicit con-
structions are known (cf. [14], [84]). Generally, \( M_\mathcal{E} \) 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), \( P \in M_\mathcal{E} \) is an extreme point
iff \( F = \{ \Sigma_J \pi_J \circ f_J : f_J \in \mathcal{G}^1(P_J) \} \) is dense in \( \mathcal{G}^1(P) \). For simple marginals mo-
re information is known if \( n = 2 \). Generally, \( M_\mathcal{E} \) can be empty, can be a
small (even one-point) set or can be a large set of distributions.

In applications \( M_\mathcal{E} \) describes a model for systems of \( n \) components, where for certain subsystems \( J \in \mathcal{E} \) one knows the distribu-
tions \( P_J \) "exact-
ly", 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 modelling problem. This means
that one should not only solve constructively the marginal problem, but more-
over construct submodels \( \mathcal{G} = \{ P_\theta, \theta \in \Theta \} \subseteq M_\mathcal{E} \) with the parameter \( \theta \) speci-
fying 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 \( M_\mathcal{E} \) (resp. a cor-
responding density with fixed marginals). For several distances characteri-
izations 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 propor-
tional 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 \( \mathcal{E} = \{ \{1\}, \ldots, \{n\} \} \) of simple marginals. In this case we use the notation
FRÉCHET BOUNDS AND THEIR APPLICATIONS

(2) \( M_\varphi = M(P_1, \ldots, P_n) \).

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 Fréchet-bounds, defined by:

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

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

(4) \( E(x) \leq F(x) \leq F(x), \quad x \in \mathbb{R} \),

where \( E(x) = (\sum_{i=1}^n F_i(x) - (n-1))_+ \), \( F(x) = \min (F_i(x))_+ \). \( E, F \) are the "lower" resp. "upper" Fréchet-bounds. \( F \) is a distribution function (is an element of the Fréchet-class \( \mathfrak{B}(P_1, \ldots, P_n) \) \), \( E \in \mathfrak{B}(P_1, \ldots, P_n) \) if \( n = 2 \), but Dall'Aglio (1972) showed that for \( n \geq 3 \), \( F \) is a df only in very exceptional cases. Based on (4) many authors established sharp bounds for \( n = 2 \) and \( \varphi(x,y) = \psi(x-y) \), \( \psi \) convex (or concave); in particular \( \varphi(x,y) = |x-y|^\alpha, \quad \alpha \geq 1 \), 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) \( U(\varphi) = \inf \{ \sum_{j \in E} \int f_j \, dP_j ; \sum_{j \in E} f_j \circ \pi_j \geq \varphi \} \),

\( L(\varphi) = \sup \{ \sum_{j \in E} \int f_j \, dP_j ; \sum_{j \in E} f_j \circ \pi_j \leq \varphi \} \),

then, obviously,

(6) \( M_\varphi (\varphi) \leq U(\varphi), \quad L(\varphi) \leq m_\varphi (\varphi) \)

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 $\varphi \leq \sum_{J \in \mathcal{E}} f_J \circ \pi_J$ 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 well-known duality theorems for (topological) vector spaces. This leads to general duality results, where $M_\mathcal{E}(\varphi), m_\mathcal{E}(\varphi)$ are replaced by

$$\widetilde{M}_\mathcal{E}(\varphi) = \sup \{ \int \varphi dP; P \in \widetilde{M}_\mathcal{E} \},$$

where $\widetilde{M}_\mathcal{E} = ba (P_J, J \in \mathcal{E})$ is the set of finite additive contents with marginals $P_J$. In a second step one has to establish conditions on $\varphi$, resp. the topology, to ensure that

$$\widetilde{M}_\mathcal{E}(\varphi) = M_\mathcal{E}(\varphi), \quad \widetilde{m}_\mathcal{E}(\varphi) = m_\mathcal{E}(\varphi)$$

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 $(X, \mathcal{B})$ (which in this section is not necessarily a product space) $\mathcal{B}_i \subseteq \mathcal{B}, i \in I$, be a system of sub-$\sigma$-algebras with probability measures $P_i \in M^1(X, \mathcal{B}_i), i \in I$. Define
\[ M = \{ P \in M(X, \mathcal{B}); P|_{\mathcal{B}_i} = P_i, i \in I \} \]

\[ \tilde{M} = \{ P \in ba(X, \mathcal{B}); P|_{\mathcal{B}_i} = P_i, i \in I \}; \]

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

\[ A \in \mathcal{B}_{i_1} \cap \mathcal{B}_{i_2} \] implies that \( P_{i_1}(A) = P_{i_2}(A) \).

Furthermore, we define

\[ F = \{ \sum_{i \in I_0} f_i; I \subset I \text{ finite}, f_i \in \mathcal{B}_i^1, P_i \} = \oplus_{i \in I} \mathcal{B}_i^1, P \]

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

\[ \mathfrak{B}^m = (\varphi \in \mathfrak{B}(X, \mathcal{B}); \exists f \in F \text{ with } \varphi \leq f), \]

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

\[ T: F \to \mathbb{R}, T(\sum_{i \in I_0} f_i) = \sum_{i \in I_0} \int f_i \, dP_i \]

is well defined.

**Theorem 1.** a) (Marginal Problem)

\( \tilde{M} \neq \emptyset \) iff \( T \geq 0 \) (i.e. \( f \in F, f \geq 0 \) implies \( Tf \geq 0 \)).

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

\[ \tilde{M}(\varphi) = \sup \{ \int \varphi \, dP; P \in \tilde{M} \} = U(\varphi) = \inf \{ Th; h \in F, \varphi \leq h \}. \]

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

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

By Hahn–Banach there exists an extension \( S \) of \( T \) to \( \mathfrak{B}^m \), \( S \leq U \). Riesz' representation theorem ensures the existence of an element \( P \in ba(X, \mathcal{B}) \) 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 \( \mu \) \( \widetilde{\mathcal{M}} \) = \( \{ P \in \text{ba}(X, \mathcal{B}) ; P|\mathcal{B}_i = P_i, i \in I, P \leq \mu \} \). Replace the operator from (14) by \( U_{\mu}(\varphi) = \inf \{ U(\varphi_0) + \int h \, d\mu ; \varphi_0 + h \geq \varphi \} \). Then the existence and duality results analogously to (14) are valid (cf. [57]).

Consider next the following assumptions:

A.1 \((X, \mathcal{B}, P), i \in I\), are compactly approximable, i.e. there exist compact set-systems \( E_i \subseteq \mathcal{B}_i \) with \( P_i(E_i) = \sup \{ P_1(E_i) ; E_i \subseteq \mathcal{B}_i, E_i \in E_i \}, i \in I \).

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

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

Theorem 2. a) If A.1 holds, then: \( M \neq \emptyset \) iff \( T \geq 0 \). Furthermore, \( M(\varphi) = U(\varphi) \) for \( \varphi \in \mathfrak{B}^1(X, \mathcal{K} \left( \bigcup_{i \in I} \mathcal{X}_i \right), P) \).

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 \( P \in \widetilde{\mathcal{M}} \) is compactly approximable on \( \mathcal{K} \left( \bigcup_{i \in I} \mathcal{X}_i \right) \) and, therefore, \( \sigma \)-additive on \( \mathcal{K} \left( \bigcup_{i \in I} \mathcal{X}_i \right) \), implying the existence of a \( \sigma \)-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 \( \sigma \)-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. \( M_{\varphi} = M(P_{1}, ..., P_{n}) \)

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 \( M_{\varphi}, U \) resp. \( M_{\varphi}. I \). These continuity properties combined with Choquet's capacity theorem yield in particular the following duality theorem. \( (E_i, \mathcal{X}_i) \) are assumed to be Hausdorff topological spaces with Borel \( \sigma \)-algebras \( \mathcal{X}_i \). This assumption is made throughout the rest of this paper. Also we assume generally that \( P_j \) are Radon measures, \( J \in \mathcal{E} \).
Theorem 3. (Kellerer (1984), Theorem 2.21)

The duality theorem \( M_\varphi(\varphi) = U(\varphi) \) is true for

a) \( \varphi \in \mathcal{B} \), the class of upper-semicontinuous functions with values in \( \mathbb{R} \). In this case there exists a maximal measure,

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

c) \( \varphi \in S_m(E) \), the class of lower majorized Suslin functions w.r.t. \( \mathcal{B} \).

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 \( B \in \mathcal{A}_1 \otimes \mathcal{A}_2 \), then

\[
M_\varphi(B) = \inf \{ P_1(B_1) + P_2(B_2); B \subseteq \bigcup_{i=1}^2 \pi_i^{-1}(B_i) \}
\]

\[
m_\varphi(B) = \sup \{ P_1(B_1) + P_2(B_2) - 1; B \supseteq B_1 \times B_2 \}.
\]

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

\[
R(Y) = \{(x, y) \in Y \times Y; \ x \leq y \}
\]

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)

\[
M_\varphi(R(Y)) = 1 - \sup \{ P_1(A) - P_2(A); A \text{ closed, isotone} \}.
\]
b) (Strassen representation theorem, cf. [101])

\[ P_1 \preceq_{st} P_2 \iff \exists P \in \text{M}(P_1, P_2) \text{ with } P(\mathbb{Y}) = 1. \]  

Strassen’s a.s. representation theorem has been very influential for the theory and applications 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 \( A = A_1 \times \ldots \times A_n \in \mathbb{X}_1 \otimes \ldots \otimes \mathbb{X}_n \) there is the obvious generalization of the Fréchet-bounds in (4).

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

\[
M_{\varnothing}(A_1 \times \ldots \times A_n) = \min \{P_i(A_i); 1 \leq i \leq n\}
\]

\[
m_{\varnothing}(A_1 \times \ldots \times A_n) = (\sum_{i=1}^{n} P_i(A_i) - (n-1))_+.
\]

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

\[
M_{\varnothing}(A_1 \times A_2) = \inf \{P_1(B_1) + P_2(B_2); A_1 \times A_2 \subseteq B_1 \times E_2 \cup E_1 \times B_2\}
\]

\[
= \min \{P_1(A_1), P_2(A_2)\}.
\]

The case \( n \geq 2 \) is proved by induction. Let \( Q_0 \in \text{M}(P_1, \ldots, P_n) \) satisfy

\[
Q_0(A_1 \times \ldots \times A_n) = \min \{P_i(A_i)\}, \quad \text{then } \sup \{P(A_1 \times \ldots \times A_{n+1}); P \in \text{M}(P_1, \ldots, P_{n+1})\} = \sup_{Q \in \text{M}(P_1, \ldots, P_n)} \sup_{P \in \text{M}(Q, P_{n+1})} P(A_1 \times \ldots \times A_{n+1})
\]

\[
= \min \{Q_0(A_1 \times \ldots \times A_n), P_{n+1}(A_{n+1})\} \quad \text{(from the case } n = 2)\]

\[
= \min \{P_i(A_i)\} \text{ by induction.}
\]

Since the opposite 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

\[ B_i := E_1 \times \ldots \times A_i \times \ldots \times E_n, \quad p_i = P_i(A_i), \quad A_1 \times \ldots \times A_n = \bigcap_{i=1}^{n} B_i. \]

The Bonferoni bounds are defined by
(19) \[ \hat{U}((p_i)) = \sup \{ P(\bigcap_{i=1}^{n} B_i); B_i \in \mathcal{A}, P(B_i) = p_i, 1 \leq i \leq n \} \]
\[ \underline{L}((p_i)) = \inf \{ P(\bigcap_{i=1}^{n} B_i); B_i \in \mathcal{A}, P(B_i) = p_i, 1 \leq i \leq n \}. \]

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

(20) \[ P \leq_s Q \text{ if } P([x, \infty)) \leq Q([x, \infty)) \]
for all \( x \in \mathbb{R}^n \) it has been proved that

(21) \[ P \leq_s Q \text{ iff } \int \varphi \, dP \leq \int \varphi \, dQ \]
for \( \Delta \)-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 \( n \geq 2 \)). (21) combined with (18), (4) imply sharp results for \( M_\varphi (\varphi) \). These are related to rearrangement inequalities (cf. [112], [81]). The case \( n = 2 \), \( \varphi(x,y) = -\varphi(x-y), \varphi \) convex, is due to Bertino (1966). Some partial results are in [79] for the lower bound \( m_\varphi (\varphi) \). An open problem is e.g. to determine \( m_\varphi (\varphi) \) for \( Y = [0,1], P_i = R(0,1) \), the uniform distribution \( 1 \leq i \leq 3 \) and \( \varphi(x) = \prod_{i=1}^{3} x_i \). \( \Box \)

2.2.3. Representation of Minimal Metrics

Some of the well-known probability metrics have a representation as a "minimal metric".

2.2.3.1. Levy-Prohorov-Metric

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

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

From (15) we obtain

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

Let \( P_1, P_2 \in M^1(Y, \mathcal{B}), \varepsilon > 0 \);

a) \[ \delta \geq 0, \text{ There exists } P \in M(P_1, P_2) \text{ with } P((x,y) \in Y \times Y; d(x,y) > \varepsilon) < \delta \]
\[ \forall A \in \mathcal{B} = \mathcal{B}(Y): P_1(A) \leq P_2(A^\varepsilon) + \delta; \]

b) \[ \delta \geq 0, \text{ There exists } P \in M(P_1, P_2) \text{ with } P(d(x,y) > \varepsilon) \leq \delta \]
\[ \Rightarrow P_1(A) \leq P_2(A^\varepsilon) + \delta, \forall A \in \mathcal{B}(Y). \] \( \Box \)
Theorem 7 implies the Strassen-representation of the Levy-Prohorov metric

\[(23) \quad \pi(P_1, P_2) = \inf \{\varepsilon > 0: P_1(A) \leq P_2(A^\varepsilon) + \varepsilon, \forall A \in \mathcal{F}(Y)\}.\]

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

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

and consider the corresponding minimal metric

\[(25) \quad \hat{K}(P_1, P_2) = \inf \{K(P); P \in \mathcal{M}(P_1, P_2)\} .\]

\[\textbf{Theorem 8.} \ (\text{Strassen (1964), Dudley (1968))}\]

\[(26) \quad \hat{K} = \pi. \quad \square\]

\(\pi\) 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.

\[(27) \quad P_n \stackrel{\mathcal{D}}{\longrightarrow} P \quad \text{if and only if} \quad \pi(P_n, P) \to 0.\]

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

\[\textbf{Theorem 9.} \ (\text{Almost sure representation theorem})\]

\(a)\)

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

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

\(b)\)


If \(P_n, Q_n \in \mathcal{M}^1(Y, d)\) 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 \((\Omega, \mathcal{A}, R)\) with \(R^n = P_n, R^X = Q_n\). \(\square\)

\[\textbf{Remark.} \ \text{If} \ P_n \in \mathcal{M}(Y), \ Y = Y_1 \times Y_2 \ \text{a product space,} \ P_n \in \mathcal{M}(Q_n, R) \ \text{and} \ P_n \stackrel{\mathcal{D}}{\longrightarrow} P, \ \text{then the following sharpening of Theorem 8 is not true:} \]

"There exist versions of \(P_n\) of the form \((X_n, Z)\), such that \((X_n)\) is a.s. convergent (cf. [70])." \(\square\)

\[\textbf{2.2.2.8 - Metrics}\]

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

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

\[
\mathcal{G}^\infty(P_1, P_2) := \inf \{ \mathcal{G}^\infty(P); P \in \mathcal{M}(P_1, P_2) \} \\
= \inf \{ \varepsilon > 0; P_1(A) \leq P_2(A^\varepsilon), \forall A \in \mathcal{G}(Y) \}.
\]

For the \( \mathcal{G}^P \)-distance, \( 1 \leq p < \infty \)

\[
\mathcal{G}^P(P) = \int d^P(x,y) dP(x)
\]

(the corresponding probability metric is \( d^P(P) = (\mathcal{G}^P(P))^{1/p} \)) the duality theorem of Section 2.2 implies for \( P_1, P_2 \in \mathcal{M}(Y) \) with \( \int d^P(x,a) dP_1(x) < \infty \)

\[
\mathcal{G}^P(P_1, P_2) = \sup \{ \int f dP_1 + \int g dP_2; f \in \mathcal{G}(P_1), g \in \mathcal{G}(P_2), f(x) + g(y) \leq d^P(x,y) \}
\]

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

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

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

\[
\mathcal{G}^1(P_1, P_2) = \sup \{ \int f d(P_1 - P_2); f \in \text{Lip}(Y) \}, \text{ where Lip}(Y) = \{ f: Y \rightarrow \mathbb{R}; |f(x) - f(y)| \leq d(x,y) \}.
\]

**Proof.** From (31)

\[
\mathcal{G}^1(P_1, P_2) = \sup \{ \int f_1 dP_1 - \int f_2 dP_2; f_1(x) - f_2(y) \leq d(x,y), f_i \in \mathcal{G}^1(P_i) \}.
\]

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

\[
f(x) = \left\{ \begin{array}{ll}
\sup \{ f_1(x_1) - d(x_1, x); x_1 \in A_1 \} & \text{if } x \in A_2 \\
\inf \{ f_2(x_2) + d(x_2, x); x_2 \in A_2 \} & \text{if } x \in A_1
\end{array} \right.
\]

Then \( f \in \text{Lip}(Y) \) and \( f \) is better then \( f_1, f_2 \), i.e. \( f(x) \geq f_1(x) \), \( x \in A_1 \), \( f(x) \leq f_2(x) \), \( x \in A_2 \), and, therefore, \( \int f_1 dP_1 - \int f_2 dP_2 = \int f_1 dR_1 - \int f_2 dR_2 = \int f d(P_1 - P_2) \) and from (31) \( \mathcal{G}^1(P_1, P_2) = \text{l}(P_1, P_2) = \sup \{ \int f d(P_1 - P_2); f \in \text{Lip}(Y) \} \).

If \( R \neq 0, R_i \neq 0 \), then the result follows from the following relations:

\[
\bar{G}^1(P_1, P_2) > \text{l}(P_1, P_2) = \text{l}(R_1, R_2) = \bar{G}^1(R_1, R_2) := \inf \{ \int d(x, y) dR(x,y), \bar{R} \in \mathcal{M}(R_1, R_2) \}
\]

\[
> \inf \{ \int d(x, y) dP(x, y); P \in \mathcal{M}(P_1, P_2) \} = \bar{G}^1(P_1, P_2).
\]

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, \( P_i = Q + \tilde{R} \in M(P_1, P_2) \) and \( \int d(x, y) dP = \int d(x, y) d\tilde{R} \). If \( R_1 = 0 \), 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 \( Y = R^1 \) the minimal \( \mathcal{G}^p \) metrics are explicitly known (cf. Gini [36], Salvemini [97], Dall’Aglio [11] - [14], Fréchet [30] - [33], Hoeffding [40], Vallender [109]).

\[
\mathcal{G}^1(P_1, P_2) = \int |F_1(x) - F_2(x)| dx \\
\mathcal{G}^p(P_1, P_2) = \int_0^1 |F_1^{-1}(t) - F_2^{-1}(t)|^p dt, \quad p \geq 1,
\]

where \( F_i \) are the df’s of \( P_i \).

For \( Y = R^k \) there are few explicit solutions. Let \( |x| \) denote the euclidean norm in \( R^k \).

**Theorem 12.** a) (Knott and Smith (1984.), Rüschendorf and Rachev (1990)) If \( \int |x|^2 dP_i(x) < \infty, \quad i = 1, 2 \), then random variables \( X, Y \) with distributions \( P_1, P_2 \) satisfy: \( E|X - Y|^2 = \mathcal{G}^2(P_1, P_2) \leq 3f: R^k \to R^1 \) 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 \( \Sigma_1, \Sigma_2 \) and \( a_1, a_2 \in R^k \):

\[
\mathcal{G}^2(N(a_1, \Sigma_1), N(a_2, \Sigma_2)) = |a_1 - a_2|^2 + \text{tr} \Sigma_1 + \text{tr} \Sigma_2 - 2 \text{tr}(\Sigma_1^{1/2} \Sigma_2^{1/2})^{1/2}.
\]

**Remark.**

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

\[
\Phi(x) = -\nabla f(x) \quad \text{is monotone}, \quad \langle x - y, \Phi(x) - \Phi(y) \rangle \geq 0, \quad \forall x, y
\]

(cf. [74], p. 99). Therefore, if \( X \) has distribution \( P_1 \), if \( \Phi \) 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. \( \mathcal{G}^2 \)-distance if and only if \( \Phi \) is monotone. Let \( \omega = \sum_{i=1}^k \Phi_i dx_i \), then \( d\omega = \sum_{i<j} \left( \frac{\partial \Phi_i}{\partial x_j} - \frac{\partial \Phi_j}{\partial x_i} \right) dx_i \wedge x_j \) and by Poincaré’s lemma we obtain: If \( \Phi \) is continuously differentiable, then \( (X, \Phi(X)) \) is an optimal coupling w.r.t. \( \mathcal{G}^2 \)-distance, if and only if
\[ \frac{\partial \Phi}{\partial x_i} = \frac{\partial \Phi}{\partial x_j}, \quad \forall i \neq j \quad \text{and} \]

\[ \Phi \text{ is monotone.} \]

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 \( \Phi(x) = \Sigma_1^{-1/2} (\Sigma_1^{1/2} \Sigma_2^{1/2} \Sigma_1^{1/2})^{-1} \Sigma_1^{-1/2} x \) 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 \( \mathcal{G}^p \)-metrics are special instances of the Monge-Kantorovich 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 \( \mathcal{G}^2 \)-metrics and their application to approximation problems are discussed in [86].

2.2.4. **Probability on Diagonals**

For \( n \geq 2 \), \( E_1 = \ldots = E_n = Y \) and for \( A \in \mathcal{A}_1 \) let

\[ \Delta_n(A) := \{(x, \ldots, x); \ x \in A \} \in \mathcal{G}^n \mathcal{A}_1 \]

and let

\[ P_1 \wedge \ldots \wedge P_n(A) := \inf \{ \sum_{i=1}^n P_i(A_i); \ A_i \in \mathcal{A}_1, \ \sum_{i=1}^n A_i = A \}. \]

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

\[ P^*(\Delta_n(A)) = \mathcal{M}_e(\Delta_n(A)) = P_1 \wedge \ldots \wedge P_n(A). \]

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

If \( n \geq 2 \), then take \( Q \in \mathcal{M}(P_1, \ldots, P_{n-1}) \) with \( Q(\Delta_{n-1}(A)) = P_1 \wedge \ldots \wedge P_{n-1}(A), \ A \in \mathcal{A}_1 \) (induction hypothesis). Then \( \mathcal{M}_e(\Delta_n(A)) = \sup_{P, Q \in \mathcal{M}(P_1, \ldots, P_{n-1})} P(\Delta_n(A)) \geq \sup_{P \in \mathcal{M}(Q, P_n)} P(\Delta_n(A)) = \inf \{ Q(B_1) + P_n(B_2); \ A_n(A) \subset B_1 \times Y \cup Y \times B_2 \}
\]

\[ = \inf \{ Q(\Delta_n(A_n)) + P_n(A \backslash A_n); \ A_n \subset A \} = \inf \{ P_1 \wedge \ldots \wedge P_{n-1}(A_1) + P_n(A \backslash A_1); \ A_1 \subset A \} = P_1 \wedge \ldots \wedge P_n(A). \] The opposite inequality is trivial. \( \Box \)
In the case $n = 2$, $A = Y$, (39) implies with $\Delta_2 = \Delta_2(Y)$ and
\begin{equation}
\phi_v(P_1, P_2) = \sup \{ P_1(B) - P_2(B); B \in \mathcal{F}_1 \},
\end{equation}
the wellknown representation of the sup-metric $\phi_v$:
\begin{equation}
\phi_v(P_1, P_2) = 1 - M_{\phi} (\Delta_2) = m_{\phi} (\Delta_2)
\end{equation}
due to Dobrushin (1969).

\subsection{Random Variables With Maximum Sums}

**Problem:** For $P_i \in M(\mathbb{R})$ with df's $F_i$, $1 \leq i \leq n$, determine the maximum resp. minimum probability of
\begin{equation}
A_n(t) = \{ x \in \mathbb{R}^n; \sum_{i=1}^n x_i \leq t \}.
\end{equation}
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
\begin{equation}
M_{\phi} (A_2(t)) = F_1 \wedge F_2 (t) = \inf_x (F_1(x) + F_2(t-x))
\end{equation}
the infimal convolution.
\begin{equation}
m_{\phi} (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))
\end{equation}

For $n > 2$ there are some particular results in [77], obtained by explicit solution of the dual problem. If e.g. $P_1 = \ldots = P_n = R(0,1)$, then
\begin{equation}
M_{\phi} (A_n(t)) = \frac{2}{n} t, \quad 0 \leq t \leq \frac{n}{2},
\end{equation}
\begin{equation}
m_{\phi} (A_n(t)) = \min \{ (\frac{2}{n} t - 1), 1 \}, \quad t > 0.
\end{equation}

If $P_i = \mathcal{B}(1, \theta)$, $1 \leq i \leq n$, then
\begin{equation}
M_{\phi} (A_n(k)) = \frac{n}{n-k} (1- \theta), \quad k < n\theta,
\end{equation}
the solution $P^* \in M_{\phi}$ being a mixture of the uniform distribution on $
\{ x: \sum_{i=1}^n x_i = k \}$ and a one point measure in $(1, \ldots, 1)$.

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

\subsection{Monte-Carlo-Simulation}

**Problem:** For $P_i \in M(\mathbb{R}^1)$ construct rv's $X_i^* \sim P_i$ with
\begin{equation}
\text{Var} \left( \sum_{i=1}^n X_i^* \right) \leq \text{Var} \left( \sum_{i=1}^n X_i \right) \text{ for any } X_i \sim P_i.
\end{equation}
For \( n = 2 \) a solution is the well-known method of "antithetic variates" (cf. Hammersley Handscomb (1964)). For \( n \geq 2 \) there are some particular results.

1. If \( P_i = \mathbb{R}(0,1) \), 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 \((1,...,n)\), 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 \( P_i = \mathbb{B}(1,\theta) \), then one can again construct \( X^*_i \sim P_i \) with \( \sum_{i=1}^{n} X^*_i \in \{k,k+1\} \), \( \frac{k}{n} \leq \theta \leq \frac{k+1}{n} \), which solve (48). The minimal value of the variance equals the cyclic function

\[
\nu_k(\theta) = a(k,\theta)(1-a(k,\theta), \quad a(k,\theta) = k\theta \pmod{1}
\]

(cf. Snijders (1984)).

In these examples it is possible to concentrate the distribution of \( \Sigma X^*_i \) "close" to \( nEX^*_1 \). For a symmetric distribution (like \( N(a,a^2) \)) and \( n = 2m \) one can choose rv's \( X_i \) with \( \sum_{i=1}^{n} X_i = nEX_1 \).

For \( P_i \in M^1(\mathbb{R}^k, \mathbb{B}^k) \), \( 1 \leq i \leq n \), 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: \( P_1 = ... = P_n \)) with error \( E|\frac{1}{n} \sum_{i=1}^{n} X_i - a|^2 \).

The corresponding problem is to determine the minimum of \( \sum_{i<j} E<X_i,X_j> \).

For \( n = 2 \) we obtain a characterization of a solution from Theorem 12 (cf. also Section 3.1): \( E<X_1^*,X_2^*> = \min_{X_1 \sim P_1, X_2 \sim P_2} E<X_1,X_2> \)

\( \Rightarrow \exists f: \mathbb{R} \to \mathbb{R} \) closed, convex, such that \( X^*_2 \in af(-X^*_1) \).

### 2.2.7. Maximally Dependent Random Variables

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

\[
\max_{1 \leq i \leq n} X^*_i \leq_{st} \max_{1 \leq i \leq n} X^*_i, \quad \forall n \in \mathbb{N}.
\]

\( X^* = (X_i) \) is called maximally dependent sequence. In the case \( P_i = \mathbb{R}(0,1) \) there is a nice geometric construction (cf. also [76]). In terms of limit theorems Lai and Robbins established that \( \max X^*_i \) 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
\[ P(\max_{1 \leq i \leq n} X_i < t) = \left( \sum_{i=1}^{n} F_{P_i}(t) - (n-1) \right)_+. \]
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 \( \mathcal{A} \)-preserving transformation \( g: [0,1] \to [0,1] \) define \( P_g \) to be the
distribution of \( (U, g(U)) \). \( P_g \) to be the distribution of \( (g(U), U) \). If \( g \) is one to one
then \( P, P_g \) are called permutation measures since \( P_g (A \times B) = \mathcal{A}^1(A \cap g^{-1}(B)), \)
\( A, B \in \mathcal{A}[0,1] \) and \( g = P \).

The only monotonic transformations of \([0,1]\) which are \( \mathcal{A} \)-preserving
are \( g_1(u) = u[\mathcal{A}], g_2(u) = 1 - u[\mathcal{A}] \), the corresponding permutation measures
are the Fréchet-distributions. The property of two random variables
\( X, Y \) that \( Y = g(X) \), \( g \) \( \mathcal{A} \)-preserving was introduced by Lancaster (1963)
under the notation: \( Y \) is completely dependent 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 \( f(x) \cdot g(y) \in L^1(\mathcal{A}^2) \). 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 mea-
surable separable metric spaces). Let \( h_i: E_i \to W_i \) be measurable, \( E_i, W_i \)
Borel spaces, \( 1 \leq i \leq n \), let \( h_j = (h_{ij}) \), \( 1 \leq j \leq J \), \( J \subseteq \{1, \ldots, n\} \), \( h = (h_1, \ldots, h_n) \).

Theorem 15. (cf. [93])

If \( \mathcal{G} \) is decomposable, then

\[
M_{\mathcal{G}}(P^h; P \in M_{\mathcal{G}}) = M(P_{\mathcal{G}}^{h_j}; J \in \mathcal{G}). \]

For the special case \( M(P_1, \ldots, P_n)^{h} = M(P_1^{h_1}, \ldots, P_n^{h_n}) \) cf. also Rachev and
Rüscheidorf (1986), Scarsini (1989). If \( h_i: ([0,1], \mathcal{A}) \to (E_i, \mathcal{A}_i, P_i) \) with \( \mathcal{A}_i^{h_i} = P_i \).
then any $P \in \mathcal{M}(P_1, \ldots, P_n)$ has a representation $P = Q^h$, $Q \in \mathcal{M}(R(0,1), \ldots, R(0,1))$.

If $E_i = \mathbb{R}$, $h_i(x_i) = F_i(x_i)$, $x_i \in (0,1)$, where $F_i$ are the df's of $P_i$, then $h_i^h = P_i$ and $P = Q^h$. Therefore,

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

$F_Q$ is the so called "copula".

(51) implies in particular for the case of simple marginals and $h_i: E_i \to \mathbb{R}$, $h_i \in \mathcal{B}_i(P_i)$:

$$M_{\mathcal{E}}(\prod_{i=1}^n h_i) = \int_0^1 \prod_{i=1}^n F_i^{-1}(u) du,$$

where $F_i^{-1}$ is the df of $P_i$.

For some decomposable cases in [14], [96] sharp bounds have been proved as e.g. for star-configurations $\mathcal{E}_i = \{(1,i), 2 \leq j \leq n\}$ or simple series-configurations $\mathcal{E} = \{(1,2), (2,3)\}$. In [96] is a discussion of two principles of deriving bounds, the method of Bonferroni-type bounds and the method of conditioning.

In the nonregular case the set $M_{\mathcal{E}}$ 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 $M_{\mathcal{E}}$, $U$ in the nonregular case seem to be strictly weaker than in the regular case. But these properties need a more detailed investigation.

**Example.** Let $n = 3$, $\mathcal{E} = \{(1,2), (2,3), (1,3)\}$, the simplest nonregular case, and let $E_i = [0,1]$. If $P_{ij} = P^{(U_{ij}, 1-U_{ij})}$ for all $i,j$, where $P^U = R(0,1)$ is uniform on $(0,1)$, then $M_{\mathcal{E}} = \varnothing$.

If $P \in \mathcal{M}([0,1]^3)$ with marginals $(P_{ij})$, $i,j \leq 3$, and $P(x: \Sigma x_i = c) = 1$, then $M_{\mathcal{E}} = \{P\}$. For the proof note that for any $Q \in M_{\mathcal{E}}$ we have

$$\int \prod_{j=1}^3 \mathbb{1}_{x_j < c} dQ(x) = \int (\Sigma x_i - c)^\mathcal{E} dP(x) = 0 \text{ i.e. } Q(x: \Sigma x_i = c) = 1.$$

This implies that the conditional distributions $P_{\mathcal{E}}|_{\mathcal{E} = \{\pi_1 = x_1, \pi_2 = x_2\} = \pi_3} = (\pi_1 = x_1, \pi_2 = x_2)$ and, therefore, $P = Q$.

If $P_{ij} = R(0,1) \otimes R(0,1)$ for all $i,j$, then $M_{\mathcal{E}_{[0,1]^3}} = R(0,1) \otimes R(0,1) \otimes R(0,1) \in M_{\mathcal{E}}$. Let $\nu_i: [0,1] \to [-1,1]$ satisfy $\int \nu_i(x) dx = 0$, $1 \leq i \leq 3$. The measures $P_{\mathcal{E}} = (1 \otimes \prod_{i=1}^3 \mathbb{1}_{x_i}) \otimes (1 \otimes \prod_{i=1}^3 \mathbb{1}_{x_i})$, $v = (v_i)$, all have two dimensional marginals $M_{\mathcal{E}_{[0,1]^3}}$, i.e. $P_{\mathcal{E}} \in M_{\mathcal{E}}$. One can explicitly construct all elements of $M_{\mathcal{E}}$, which are continuous w.r.t. $M_{\mathcal{E}_{[0,1]^3}}$ (cf. [87]).
For \( P_{ij} \in M(E_i \times E_j) \) let \( P_{ij} |_{\{\pi_i = \pi_j\}} \) be the conditional distribution and define for \( A \in \mathcal{A}_1 \otimes \mathcal{A}_3 \)

\[
U_{13 | x_2}(A) := \inf \{ P_{1 | x_2}(A_1) + P_{3 | x_2}(A_2) \mid A \subset A_1 \times E_3 \cup E_1 \times A_2 \}
\]

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

**Theorem 16.** (cf. [96])

a) If \( \mathcal{G} = \{(1,2), (2,3)\}, B \in \mathcal{B}_1 \otimes \mathcal{B}_2 \otimes \mathcal{B}_3 \), then

\[
M_{\mathcal{G}}(B) = \int U_{13 | x_2}(B_{x_2}) dP_2(x_2)
\]

\[
m_{\mathcal{G}}(B) = \int L_{13 | x_2}(B_{x_2}) dP_2(x_2).
\]

b) If \( \mathcal{G} = \{(1,2), (1,3), (2,3)\} \), then

\[
M_{\mathcal{G}}(\phi) \leq U(\phi) := \min \{ \int U_{23 | x_1}(\varphi_{x_1}) dP_1(x_1), \int U_{13, x_2}(\varphi_{x_2}) dP_2(x_2), \int U_{12 | x_3}(\varphi_{x_3}) dP_3(x_3) \}
\]

for \( \phi \in \mathbb{R}_m(E) \), 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

\[
M_{\mathcal{G}}(A) \leq \bar{U}(1_A) \leq \min \{ P_{ij}(A_1 \times A_j) \}.
\]

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 \( \mathcal{G} = \mathcal{J}_2^3 = \{(1,2), (1,3), (2,3)\} \) let

\[
C(P_{12}, P_{23}) = \{P_{13} \mid M(P_{12}, P_{13}, P_{23}) \neq \emptyset \}
\]

be the compatibility set of \( P_{12}, P_{23} \). Dall'Aglio (1959, 1972) proved that in the case \( E_1 = \mathbb{R}^1 \):

\[
F_{13}(x_1, x_3) := \int \max \{ F_{1 | x_2}(x_1) + F_{2 | x_2}(x_3) - 1, 0 \} dP_2(x_2)
\]

\[
\leq F_{13}(x_1, x_3) \leq F_{13}(x_1, x_3) = \int \min \{ F_{1 | x_2}(x_1), F_{2 | x_2}(x_3) \} dF_2(x_2),
\]

\( F_{13}, F_{13} \) 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 \( \mathcal{G} = \mathcal{J}_2^3 \).

**Theorem 17.** (cf. [96])

\( P_{13} \in C(P_{12}, P_{23}) \iff \forall \varphi = \varphi(x_1, x_3) \) bounded, measurable:
(60) \[ \tilde{L}_{13}(\varphi) \leq \int \varphi \, dP_{13} \leq \tilde{U}_{13}(\varphi), \] where
\[ \tilde{U}_{13}(\varphi) = \int U_{13|x_2}(\varphi) \, dP_2(x_2), \quad \tilde{L}_{13}(\varphi) = \int L_{13|x_2}(\varphi) \, dP_2(x_2). \]

It is not sufficient to consider indicator functions only.

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

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

(61) \[ \mathcal{M}_c (c) = \inf \{ \int f \, dP_1 + \int g \, dP_2 ; f \in \mathfrak{B}_1(P_1), g \in \mathfrak{B}_1(P_2), c(x,y) \leq f(x) + g(y) \} \]

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

If \( (f,g) \) are admissible (i.e. \( f(x) + g(y) \geq c(x,y) \)) and \( P \in \mathcal{M}_c \), then \( (P,f,g) \) are solutions if and only if

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

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

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. \( \mathfrak{B}^2 \)-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) \[ f^*: E_2 \to \overline{\mathbb{R}}^1, f^*(y) = \sup_{x \in E_1} (c(x,y) - f(x)) \]

and the **doubly \( c \)-conjugate**

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

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 **equality domains** of \((f, f^*)\) by

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

Define the class of **c-convex** functions

\[
\Gamma^c(E_1) = \{ h; E_1 \rightarrow \mathbb{R}; h(x) = \sup_{i \in I} \left[ c(x,y_i) + a_i \right] \text{ for some } a_i \in \mathbb{R}, \quad y_i \in E_2, \quad i \in I \} \\
\Gamma^c(E_2) = \{ h; E_2 \rightarrow \mathbb{R}; h(y) = \sup_{i \in I} \left[ c(x_i,y) + b_i \right] \text{ for some } b_i \in \mathbb{R}, \quad x_i \in E_1, \quad i \in I \}
\]

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

If \( c(x,y) = \langle x,y \rangle \) \( x, y \in E \) 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 functions is very large, so that in these cases the reduction is not very interesting (cf. [19], [11]).

**Theorem 18.** For \( c \in \mathcal{B}_m \) with \( \int c(x,a) dP_i(x) < \infty, i = 1,2 \), we have: \( P \in \mathcal{M}_c \) is a maximal measure induced by random variables \( X \sim P_1, Y \sim P_2 \), if and only if

\[
Y \in E_c f(X) \text{ a.s. for some c-convex } f \in \mathcal{B}^1(P_1) \text{ or, equivalently, if and only if } X \in E_c f^*(Y).
\]

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

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

From (68) it is of interest to characterize the equality sets of c-convex functions. For \( \tilde{f}: E_1 \rightarrow \mathbb{R} \epsilon > 0 \), define the \( \epsilon \)-c-subdifferential

\[
\partial_{c, \epsilon} f(x) = \{ y; f(x') - f(x) \geq c(x', y) - c(x, y) - \epsilon, \forall x' \in E_1 \}.
\]
\( \partial_c f(x) = \partial_{c,0} f(x) \) 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])

\[
(70) \quad y \in \partial_c f(x) \iff y \in E_c f(x) \text{ (i.e. } f(x) + f^*(y) = c(x,y)\text{)}
\]

\[
\iff f(x) - c(x,y) = \inf_{x'} (f(x') - c(x',y)) \iff x \in \partial_c f(y), \text{ if } f \text{ is c-convex}.
\]

If \( \partial_{c,0} f(x) \neq \emptyset \) for all \( 0 < \varepsilon \leq \varepsilon_0 \), 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

\[
(71) \quad c(y,\Phi(x)) + c(x,\Phi(y)) \leq c(x,\Phi(x)) + c(y,\Phi(y)), \forall x,y \in A.
\]

**Proof.** Since \( f(y) - f(x) \geq c(y,\Phi(x)) - c(x,\Phi(x)) \) and \( f(x) - f(y) \geq c(x,\Phi(y)) - c(y,\Phi(y)) \), (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 \( \Phi \).

\[
(72) \quad \langle y - x,\Phi(y) - \Phi(x) \rangle \geq 0.
\]

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 \( h(y) = f(y) - f(x) - c(y,\Phi(x)) - c(x,\Phi(x)) \geq 0 \) has a minimum in \( y = x \) and, therefore,

\[
(73) \quad \nabla f(x) = \partial_1 c(x,\Phi(x)).
\]

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

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

Similarly,

\[
(75) \quad \nabla f^*(\Phi(x)) = \partial_2 c(x,\Phi(x))
\]

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

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

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

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

\[
= c(0,\Phi(0)) + \int_{0 \to x} d(c(u,\Phi(u))) = c(x,\Phi(x)).
\]
Therefore, the condition that $\Phi(x)$ is the $c$-subgradient of a differentiable function $f$, is equivalent to

$$f(x) + f^*(y) = f(x) - f(\Phi^{-1}(y)) + f(\Phi^{-1}(y)) + f^*(y) = c(\Phi^{-1}(y)) + \int_{\Phi^{-1}(y) \to y} \delta_1 c(u, \Phi(u)) \cdot du \geq c(x, y),$$

equivalently, to the differential characterization

$$\int_{\Phi^{-1}(y) \to y} [\delta_1 c(u, y) - \delta_1 c(u, \Phi(u))] \cdot du \leq 0, \quad \forall x, y.$$ (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 $\Phi$ is continuously differentiable, injective and if $\delta_1 c(x, \Phi(x)) \cdot dx, i = 1, 2$ is closed, then: $\Phi(x) \in \partial_c f(x), \quad \forall x,$ 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 $\Phi: [0, \infty) \to [0, \infty)$ be a Young-function i.e. $\Phi$ is right continuous, nondecreasing, $\Phi(0) = 0$ and $\Phi(x) \to \infty$ as $x \to \infty$ and define the generalized inverse $\Phi^{-1}(y) = \sup \{ x : \Phi(x) \leq y \}.$ The Young-inequality states the inequality:

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

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 corresponding measure $m$

$$h_1(x) = m(\{(s, t); 0 \leq s \leq x, 0 \leq t \leq \Phi(s)\}) = \int_0^x \int_0^t dF(t|s) dF_1(s),$$

$$h_2(y) = \int_0^y \int_0^t dF(s|t) dF_2(t),$$

$F(\cdot), F_1(\cdot)$ denote the conditional resp. marginal "distribution" functions.

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

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

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

$$C = \{(s, t); 0 \leq t \leq y, 0 \leq s \leq \Phi^-(y)\}.$$

Then

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

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.  

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 \( m = f x^2 \), then we obtain more explicitely:

\[
(84) \quad h_1(x) = \int_0^x \left( \int_0^x f(s, t) \, dt \right) \, ds = \int_0^x (\partial_1 F(s, \Phi(s)) - \partial_1 F(s, 0)) \, ds
\]

and

\[
(85) \quad h_2(y) = \int_0^y \partial_2 F(\Phi^-(t), t) \, dt - F(0, y) + F(0, 0),
\]

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

\[
(86) \quad F(x, y) \leq F(0, 0) + \int_0^x \partial_1 F(s, \Phi(s)) \, ds + \int_0^y \partial_2 F(\Phi^-(t), t) \, dt.
\]

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

Example. Let for \( \alpha > 1 \), \( F(x, y) = -|x - y|^{\alpha} \), \( x, y \in \mathbb{R}_+^1 \), then \( \partial_1 \partial_2 F(x, y) = \alpha(\alpha - 1)|x - y|^{\alpha - 2} \geq 0 \). Therefore, by (86) we obtain the inequality

\[
(87) \quad |x - y|^{\alpha} \geq \alpha \int_0^x |s - \Phi(s)|^{\alpha - 1} \sgn(s - \Phi(s)) \, ds
\]

\[
+ \alpha \int_0^y |t - \Phi^-(t)|^{\alpha - 1} \sgn(t - \Phi^-(t)) \, dt,
\]

where \( \sgn(x) = \begin{cases} 1 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases} \) and \( \Phi \) 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 \( \Phi(X\cdot) \preceq Y \preceq \Phi(X) \) and \( \Phi \) 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, \( F(x, y) = P(X \leq x, Y \leq y) \), then

\[
(87) \quad F(x, y) \leq f(x) + g(y), \text{ where } f(x) = \int_{-\infty}^x P(Y < \Phi(s) | X = s) \, dX(s) \quad \text{and} \quad g(y) = \int_{-\infty}^y P(X < \Phi^-(s) | Y = s) \, dP(s).
\]
Proof. \( F(x,y) = P(X \leq x, Y \leq \Phi(X) \wedge y) + P(X \leq x, Y \leq y, Y > \Phi(X)) \leq \)
\( P(X \leq x, Y \leq \Phi(X)) + P(X < \Phi(Y), Y \leq y) = \int_0^x P(Y \leq \Phi(X)|X = s)dP(Y|s) + \)
\( \int_0^y P(X > \Phi^{-1}(Y)|Y = s)dP_Y(s) = f(x) + g(y). \)

In (87) we have equality, iff
\[ \Phi(X^-) \leq Y \leq \Phi(X) \text{ a.s.} \]

(85), (87) imply optimal coupling results and Fréchet bounds for \( \Delta \)-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, \( 1 \leq i \leq n \), then:
\[ \prod_{i=1}^n f_i(t_i) \leq \sum_{i=1}^n \int_0^{t_i} \left( \prod_{j \neq i} f_j \right) df_i. \]

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

Consider the curve \( y(t) = (f_1(t), \ldots, f_n(t)) \), \( t \geq 0 \) and the points \( P_i \equiv (f_j(t_i))_{1 \leq j \leq n}, 1 \leq i \leq n \), \( A_i = (f_j(t_i))_{1 \leq j \leq n} \). Define, furthermore,
\[ V_i : = \{ x \in \mathbb{R}_+^n : x_i < f_i(t_i), x_j < f_j(f_j^{-1}(x_j)), j \neq i \} \text{ and } V_i : = [0,A_i] = [0,f_i(t_i)] \times \ldots \times [0,f_n(t_n)]. \]

**Theorem 23. (Generalized Oppenheim Inequality)**

Let \( m \) be a Radon measure on \( \mathbb{R}^n_+ \) and define \( \Phi(t_1, \ldots, t_n) = m([0,A]) \).
\[ h_i(t_i) = m(V_i), 1 \leq i \leq n, \text{ then:} \]
\[ \prod_{i=1}^n f_i(t_i) \leq \sum_{i=1}^n \int_0^{t_i} \left( \prod_{j \neq i} f_j \right) ds = \sum_{i=1}^n \int_0^{t_i} \left( \prod_{j \neq i} f_j(u) \right) df_i(u), \text{ the Oppenheim inequality.} \]

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: \( \Phi(t_1, \ldots, t_n) = F(f_1(t_1), \ldots, f_n(t_n)), \)
\[ h_i(t_1) = \mathbb{P}(U_1 \leq f_i(t_1), \ldots, U_n \leq f_j \circ f_i(U_j), j \neq i) = \int_{-\infty}^{t_1} \mathbb{P}(U_1 \leq f_j(u), j \neq i \mid U_i = f_i(u)) d\mathbb{P}_i(U_i)(u), \]
then
\[ \varphi(t_1, \ldots, t_n) \leq \sum_{i=1}^{n} h_i(t_i). \]

**Proof.** \( \varphi(t_1, \ldots, t_n) = \mathbb{P}(U_1 \leq f_i(t_i), 1 \leq i \leq n) = \mathbb{P}(f_i^-(U_i) \leq t_i, 1 \leq i \leq n) \leq \sum_{i=1}^{n} \mathbb{P}(f_i^-(U_i) \leq t_i, f_j^-(U_j) \leq f_i^-(U_i), j \neq i) \leq \sum_{i=1}^{n} \mathbb{P}(U_1 \leq f_i(t_1), \ldots, U_n \leq f_j \circ f_i^-(U_i), j \neq i) = \sum_{i=1}^{n} \int_{-\infty}^{t_i} \mathbb{P}(U_1 \leq f_j \circ f_i(U_j), j \neq i \mid U_i = s) d\mathbb{P}_i(U_i)(s). \]

\[ \sum_{i=1}^{n} t_i \leq \sum_{i=1}^{n} \int_{-\infty}^{t_i} \mathbb{P}(U_1 \leq f_j(u), j \neq i \mid U_i = f_i(u)) d\mathbb{P}_i(U_i)(u). \]

4. **Some Statistical Applications and Problems**

4.1. **Marginal Sufficiency**

Let \( \mathcal{P} \) be a dominated set of product measures on \((E, \mathcal{F}) = \prod_{i=1}^{n} (\mathbb{R}, \mathcal{B}(\mathbb{R}))\) and define \( T : (E, \mathcal{F}) \to (Y, \mathcal{B}(\mathbb{R})) \) to be **marginally sufficient** for \( \mathcal{P} \), if for all \( 1 \leq i \leq n \) and \( \varphi \in \mathcal{B}(E, \mathcal{F}) \), \( \varphi = \varphi(x_i) \), there exists \( \tilde{\varphi} \in \mathcal{B}(E, \sigma(T)) \) with
\[ \tilde{\varphi} = E_P(\varphi \mid T)[P], \forall P \in \mathcal{P}. \]
Huzurbazar proposed the following conjecture.

**Huzurbazar conjecture:**

\[ \text{Marginal sufficiency of } T \text{ implies sufficiency} \]
(i.e. (93) is true for any \( \varphi = \varphi(x_1, \ldots, x_n) \in \mathcal{B}(E, \mathcal{F}) \)).

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 \( P = \otimes P_i, Q = \otimes Q_i \in \mathcal{P} \), \( f = \prod f_i, g = \prod g_i \) be densities w.r.t. a dominating measure \( \mu = \otimes u_i \). Let \( h_i = g_i / f_i \), \( \varphi(x) = (\ell h_i(x_1), \ldots, \ell h_i(x_n)) \) and let \( P_y, Q_y \) denote the conditional distributions of \( P, Q \) given \( T = y \). If \( T \) is marginally sufficient, then \( \tilde{P}_y : = P \varphi \mid T = y = (P_y)\varphi \) and \( \tilde{Q}_y : = Q \varphi \mid T = y = (Q_y)\varphi, y \in Y \), are probability measures on \( \mathbb{R}^n \) with identical marginals. Using
\[ \frac{dQ^y}{dP^y}(x) = \frac{dP^T}{dQ^T}(y) \Pi h_i(x_i) = \frac{dP^T}{dQ^T}(y) \exp(\Sigma \ell h_i(x_i)) \] one concludes:
\[ \frac{d\tilde{Q}^y}{d\tilde{P}^y}(z) = \frac{dP^T}{dQ^T}(y) \exp(\Sigma z_i). \]
Let $U, V$ be orthogonal, nonnegative measures with $P_S - Q_S = U - V$, then

1. $U, V$ have identical marginals,

2. $U(\Sigma x_k \leq \ell) = 0, V(\Sigma x_k > \ell) = 0$ for some $\ell$ (namely $\ell = -\ell n \frac{dP^T}{dQ^T}(y)$),

3. $\int |x_k| dU < \infty, 1 \leq k \leq n$.

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

\begin{equation}
1., 2., 3. \text{ implies that } U = V = 0.
\end{equation}

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 $\Phi$.

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

\begin{equation}
P = \frac{3}{4} S_{\varepsilon_{(1.1,-1)}} + \sum_{k=2}^\infty \frac{1}{2^{k+2}} S_{\varepsilon_{(2^{k-1},1,1,1,1-2^{k-1})}}
\end{equation}

\begin{equation}
Q = \varphi P,
\end{equation}

where $S_{\varepsilon_{(1.1,-1)}} = \varepsilon_{(1,1,-1)} + \varepsilon_{(-1,1,1)} + \varepsilon_{(-1,1,1)}$, $\varepsilon_x$ the one point measure in $x$. Then the marginals of $P, Q$ are identical and equal to

\begin{equation}
\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)})
\end{equation}

and $P\{x \in \mathbb{Z}^3; \Sigma x_i = 1\} = 1$, while

\begin{equation}
Q\{\Sigma x_i = -1\} = 1.
\end{equation}

\[\square\]

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

**Theorem 25.** (Generalized Huzurbazar conjecture, cf. [95])

If $T: (E, \mathcal{A}) \to (Y, \mathcal{B})$ is partially sufficient for $\sigma(f_k), 1 \leq k \leq n$, then $T$ is sufficient for $(P_n, Q_n)$.

The proof uses the following two lemmas:

**Lemma 1.** (cf. [78], Prop. 6)

If $P_i \in M^1(\mathbb{R}^1), 1 \leq i \leq n, P, Q \in M(P_1, \ldots, P_n)$, then:

\begin{equation}
P \leq_{st} Q \text{ implies } P = Q.
\end{equation}

\[\square\]
Lemma 2. (cf. Simons (1980))

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

\begin{equation}
\Pr_{1\mathcal{B}} \leq_{st} \Pr_{Q\mathcal{B}}.
\end{equation}

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

4.2. Optimal Combination of Tests of Marginals

Let $P_i, Q_i \in M^1(E_i, \mathfrak{A}_i), 1 \leq i \leq n$, and consider the test problem with hypothesis $\Theta_0 = M(P_1, \ldots, P_n)$ 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 component 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 $\Theta_0', \Theta_1$ then to take the test for that component which allows for a certain test level $\alpha$ 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 criterion. We consider the tests of level $\alpha$

\begin{equation}
\Phi_\alpha(\Theta_0) = \{\varphi \in \Phi : \Pr_{\varphi} \leq \alpha, \forall P \in M(P_1, \ldots, P_n) = \Theta_0\}
\end{equation}

and the maximin-risk

\begin{equation}
\beta(\alpha, \Theta_0, \Theta_1) = \sup_{\varphi \in \Phi_\alpha(\Theta_0)} \inf_{P \in \Theta_1} \Pr_{\varphi}.
\end{equation}

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

\begin{equation}
d_{\mathcal{V}}(P, Q) = \sup \{P(A) - Q(A) ; A \in \mathfrak{A}\}
\end{equation}

\begin{equation}
d_{\mathcal{V}}(\mathcal{P}, \mathcal{Q}) = \inf \{d_{\mathcal{V}}(P, Q) ; P \in \mathcal{P}, Q \in \mathcal{Q}\}
\end{equation}

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

\begin{equation}
h_\alpha(x) = \alpha x + \max_{1 \leq i \leq n} d_{\mathcal{V}}(Q_i, x P_i), x \geq 0.
\end{equation}
Theorem 26. (cf. [85])

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

\[ \beta(\alpha, \Theta_o, \Theta_1) = h_\alpha(x^*). \]

a) \[ d_v(Q, x^*P) = d_v(\Theta_1, x^*\Theta_o). \]

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

\[ d_v(Q, x^*P) = d_v(\Theta_1, x^*\Theta_o). \]

then there exists a LQ-test \( \phi^* \) for \( (P, Q) \) with critical value \( x^* \)

such that \( \phi^* \) is a maximin level \( \alpha \)-test, i.e.

\[ \forall x \geq 0 \text{ holds:} \]

\[ d_v(M(Q_1, ..., Q_n), x \cdot M(P_1, ..., P_n)) = \max_{1 \leq i \leq n} d_v(Q_i, x P_i). \]

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

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

\[ d_v(Q_1, xP_1) = \max_j d_v(Q_j, xP_j), \forall x \geq 0. \]

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 \( M(P_1, ..., P_n) \) is considered as a neighborhood of \( P_1 \otimes ... \otimes P_n \), then for a test \( \phi, M_\phi(\phi) \) 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 \( \Psi = M(P_1, ..., P_n) \) for certain functions \( g : R^1 \rightarrow R^1 \).

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 \( \bar{P} : = \otimes_{i=1}^k P_i \)

and let \( D_g \) denote the unbiased estimators of \( g. \)
Theorem 27. (cf. [89])

a) \( D_\circ = F_1 = \{ \sum_{i=1}^k f_i(x_i); f_i \in G_1^1(P_i), \int f_i \, dP_i = 0, 1 \leq i \leq k \}. \)

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

\[
(108) \quad d^\ast = d - E_P(d|F_2^P)
\]

is MVU for \( g \) in \( P \), where \( F_2^P \) denotes the closure of \( F_2 = \{ \sum f_i(x_i); f_i \in G_2^2(P_i), \int f_i \, dP_i = 0 \} \) in \( G_2^2(P) \).

c) If \( d \in D_g \), then

\[
(109) \quad d^\ast = d - \sum_{i=1}^k \int d \, d \, P_i + k \int d \, d \, \overline{P}
\]

is MVU for \( g \) in \( \overline{P} \).

The projections occurring 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 \( \mathcal{P}^n = \{ P^n; P \in \mathcal{P} \} \) and the corresponding optimal estimator is given by

\[
(110) \quad d^\ast_n(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^n d^\ast(x_i).
\]

An estimator sequence for a differentiable functional \( g \), which is asymptotically optimal on the whole model or a subset \( \mathcal{P}_\circ \subset \mathcal{P} \), should have the stochastic expansion

\[
(111) \quad \sqrt{n} (d^\ast_n(x_1, \ldots, x_n) - g(P)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n g_P(x_i) + o_P(n), P \in \mathcal{P}_\circ,
\]

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

\[
(112) \quad T(P, \mathcal{P}) = (F_2^\perp)^P,
\]

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 \( g_P \in T(P, \mathcal{P}) \), 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 \( \mathcal{P}_\alpha \subset \mathcal{P}, \alpha > 0, k = 2 \), consisting of 'positive dependent' measures \( P \) with

\[
(113) \quad P(A \times B) \geq \alpha P_1(A)P_2(B), \forall A, B.
\]
References


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


